PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 38/10, 38/17, C07K 7/08, 14/81, C12N 1/15, 1/21, 5/10, 15/15

A1

(11) International Publication Number: **WO 97/26001**

(43) International Publication Date: 24 July 1997 (24.07.97)

(21) International Application Number: PCT/US97/00908

(22) International Filing Date: 16 January 1997 (16.01.97)

(30) Priority Data:
586,592 16 January 1996 (16.01.96) US

(71) Applicant: NORTHWESTERN UNIVERSITY [US/US]; 1801 Maple Avenue, Evanston, IL 60201-3135 (US).

(72) Inventors: GOLDBERG, Erwin; 2756 Central Park Avenue, Evanston, IL 60201 (US). WEINBERG, Patricia, O'Hern; Apartment 414, 1121 Church Street, Evanston, IL 60201 (US).

(74) Agents: CROOK, Wannell, M. et al.; Sheridan Ross P.C., Suite 3500, 1700 Lincoln Street, Denver, CO 80203 (US).

(81) Designated States: AU, CA, CN, IL, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: PROTEINS AND PEPTIDES FOR CONTRACEPTIVE VACCINES AND FERTILITY DIAGNOSIS

(57) Abstract

The invention comprises novel proteins and peptides derived from these proteins. The proteins are unique to sperm and testes, and the proteins and peptides are useful in vaccines for contraception in mammals. The proteins and peptides are also useful in diagnostic assays for assessing infertility. The invention also provides DNA molecules coding for the proteins and peptides and host cells containing the DNA molecules linked to expression control sequences for producing the proteins and peptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

PROTEINS AND PEPTIDES FOR CONTRACEPTIVE VACCINES AND FERILITY DIAGNOSIS

This invention was developed in part by a subcontract under grant U54 HD 29099 from the National Institutes of Health (NIH) and a grant from the Contraceptive Research and Development Program (CSA-92-099) under a Cooperative Agreement with the U.S. Agency for International Development (DPE-3044-A-00-6063-00), which in turn receives funds for AIDS research from an interagency agreement with the National Institute of Child Health and Human Development (NICHD). The U.S. government may have rights in the invention.

FIELD OF THE INVENTION

This invention relates to novel proteins and peptides and their use in contraceptive vaccines and to assess infertility. The invention also relates to DNA molecules coding for the proteins and peptides and host cells containing the DNA molecules linked to expression control sequences for producing the proteins and peptides.

BACKGROUND OF THE INVENTION

Mammalian spermatozoa are highly specialized both in structure and function. These cells are the product of a developmental program that involves the expression of genes unique to the testes and of testis-specific variants of common somatic genes. Why testis and sperm should need specialized isoforms of common proteins or genes that are expressed only during spermatogenesis remains to be established.

Idiopathic infertility is characterized clinically as the inability to achieve a pregnancy by cohabiting couples with no apparent anatomical or functional reproductive pathology. In about 10% of such cases, the cause is attributed to immunological phenomena, including circulating antisperm antibodies in one or both partners. Presumably, such antibodies target to spermatozoa and, as a consequence, conception is blocked or fails. Additionally, there is indirect evidence of an association

between infertility and antisperm antibodies in both male and female patients. With respect to the subject of immunologic infertility, see Witkin et al., Am. J. Obstet. Gynecol., 158, 59-62 (1988); Clarke et al., Fertil. Steril., 49, 1018-1025 (1988); Mathur et al., Fertil. Steril., 36, 486-495 (1981); Menge, in Immunological Aspects Of Infertility And Fertility Regulation, pages 205-224 (Dhindsa and Schumacher eds. 1981); and Isojima et al., Am. J. Obstet. Gynecol., 101, 677-683 (1968).

These observations regarding immunologic infertility led to the suggestion that a vaccine based on a sperm antigen could provide an effective and innovative contraceptive technology. A number of sperm-specific proteins and peptides have been evaluated for use in contraceptive vaccines. See generally, Alexander et al., Reprod. Fertil. Dev., 6, 273-280 (1994) and Aitken et al., Brit. Med. Bull., 49, 88-99 (1993). For a recent review of sperm antigens, see Diekman and Goldberg, in Immunology Of Human Reproduction, Chapter 1 (1995). The testis-specific isoform of lactate dehydrogenase, LDH-C_4, and peptides derived from it are perhaps the most extensively characterized sperm antigens. See U.S. Patents Nos. 4,290,944, 4,310,456, 4,353,822, 4,354,967, 4,377,516, 4,392,997, 4,578,219, 4,585,587, 4,782,136, and 4,990,496; Wheat and Goldberg, in Isozymes: Current Topics In Biological and Medical Research, Volume 7: Molecular Structure and Regulation, pages 113-130 (1983); Millan et al., Proc. Natl. Acad. Sci. USA, 84, 5311-5315 (1987); Goldberg, in Gamete Interaction: Prospects For Immuno-contraception, pages 63-73 (Alexander et al. eds. 1990); LeVan and Goldberg, Biochem. J., 273, 587-592 (1991); O'Hern and Goldberg, Proceed. Intern.: Symp. Control. Rel. Bioact. Mater., 20, 394-395 (1993); O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993); Kaumaya et al., J. Molec. Recog., 6, 81-94 (1993); and O'Hern et al., Biol. Reprod., 52, 331-339 (1995).

Even though several sperm antigens have been identified, there r mains a need to identify additional such antigens. In particular, it may be necessary to use a contraceptive vaccine containing several sperm antigens in genetically diverse populations of mammals, such as humans, to obtain effective contraception.

SUMMARY OF THE INVENTION

The invention provides purified proteins and peptides whose sequences comprise the sequence of an epitope of one of these proteins. The proteins and peptides are described in detail below.

The proteins are unique to sperm and testis, and the proteins and peptides can be used in vaccines for contraception in mammals. Accordingly, the invention further provides: (1) immunogens comprising a peptide linked to a carrier, the peptide being capable of producing an antibody that reacts specifically with one of the proteins of the invention and having a sequence comprising a sequence which forms a B-cell epitope of the protein; and (2) vaccines comprising the proteins (or immunogenic portions thereof), peptides and immunogens in a delivery system.

In addition, the proteins and peptides can be used in diagnostic assays for assessing infertility. The assays and kits for performing the assays are also part of the invention.

Finally, the invention provides DNA molecules coding for the proteins and peptides, and host cells containing the DNA molecules linked to expression control sequences, for producing the proteins and peptides.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Diagram comparing the sequences of somatic and testis-specific isoforms of calpastatin.

Figure 2: Computer-generated hydropathy plot comparing the first forty-one amino acids of somatic (solid

bars) and testis-specific (open bars) isoforms of calpastatin.

Figure 3: Western blot of human tissue extracts (lane 1 - testis, lane 2 - sperm, lane 3 - liver) probed with affinity-purified rabbit antiserum to a peptide having the sequence of a B-cell epitope found only on the testis-specific isoform of calpastatin.

Figure 4: Graph of ELISA results. In particular, absorbance at 405 nm is plotted versus weeks post primary immunization of macaques with a peptide having the sequence of a B-cell epitope found only on testis-specific isoform of calpastatin linked to a universal T-cell epitope by a four-amino acid linker.

Figure 5: Diagram of the technique of epitope mapping by nested deletions for clone C-2 and photograph of Coomasie blue-stained PAGE gel after separation of the resultant truncated proteins.

Figure 6: Western blots of truncated proteins produced by nested deletions performed to identify B-cell epitopes on the protein produced by clone C-2.

Figure 7: Diagram illustrating epitope identification for clone C-2.

Figure 8: Computer-generated plot of the occurrence of the amino acid valine along the length of the clone L-7 protein.

Figure 9: Western blots of truncated proteins produced by nested deletions performed to identify B-cell epitopes on the protein produced by clone L-7.

Figure 10: Diagram illustrating epitope identification for clone L-7.

DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

In a first aspect, the invention provides a purified protein which is a testis-specific isoform of calpastatin. "Testis-specific" is used herein to mean that the isoform is found in the testes and sperm, but is not found in other tissues. In contrast to the testis-specific isoform are

the somatic isoforms of calpastatin. The somatic isoforms are those found in one or more, generally several, types of tissues. The somatic isoforms may be found in testes and sperm but, if so, will also be found in at least one other type of tissue.

Clone Y-19, coding for a human testis-specific isoform of calpastatin, was identified by screening a human testis cDNA library with sera from infertile patients positive for antisperm antibodies (see Example 1 below). The complete sequence of this human testis-specific isoform of calpastatin is given in Chart A below.

Affinity-purified antiserum specific for this testis-specific isoform of calpastatin was used to localize th isoform on human sperm by immuno-fluorescence. Diffuse, granular fluorescence was observed throughout the acrosome, and intense fluorescence was observed in the equatorial segment of the sperm (see Example 4).

Calpastatin is the peptide inhibitor of calpain, a cysteine protease. Calpain has been localized to the sperm head and appears to be involved in the acrosome reaction. See, Schollmeyer, Biol. Reprod., 34, 721-731 (1986). Although not wishing to be bound by any particular theory, it is believed that infertility in individuals having antibodies directed to testis-specific calpastatin occurs as follows. The acrosome reaction, which must occur in order for the sperm to penetrate the zona pellucida of the egg, is triggered by an influx of Ca^{+2}. Wasserman, Annu. Rev. Cell Biol., 3, 109-142 (1987). Calpain, then, in the presence of the Ca^{+2} would hydrolyze calpastatin, thereby releasing protease inhibition and permitting proteolytic activity in membrane fusion phenomena. Goll et al., Bioessays, 14, 549-556 (1992). Perturbation of this sequence of events by antibodies directed to testis-specific calpastatin would compromise fertilization and concomitantly cause infertility. Preliminary studies hav demonstrated loss of calpastatin immunoreactivity from acrosome-reacted sperm, a result predicted from this

theory. Also, the immunofluorescence studies described above show that testis-specific calpastatin is found on the surface of sperm and would, therefore, be accessible to antibodies.

The invention further provides a protein which is the protein produced by clone C-2. Clone C-2 is a human cDNA clone that was identified by screening a human testis cDNA library with sera from infertile patients positive for antisperm antibodies (see Example 1 below). The C-2 protein is found in testis and sperm, but it is not found in other tissues. The complete amino acid sequence of the C-2 protein is set forth in Chart B below.

The invention also provides a protein which is the protein produced by clone L-7. Clone L-7 is a human cDNA clone that was identified by screening a human testis cDNA library with sera from infertile patients positive for antisperm antibodies (see Example 1 below). The L-7 protein is found in testis and sperm, but it is not found in other tissues. Affinity-purified antiserum specific for the L-7 protein was used to localize the L-7 protein on human sperm by immunofluorescence. Fluorescence was observed throughout the acrosome. The complete amino acid sequence of the L-7 protein is set forth in Chart C below.

As noted above, the Y-19, C-2 and L-7 proteins are human proteins. Corresponding proteins in other mammals would be expected to be at least 70% homologous to these human proteins. The corresponding proteins in other mammals can be obtained by the method described in Example 1 or by using the sequences given in Charts A, B and C to design DNA probes which can be used to screen testis gen libraries, preferably cDNA libraries, of other mammals. Methods of making gene (e.g., cDNA) libraries, designing probes for screening them, identifying and isolating a desired clone, producing protein from the clone, etc., are well known in the art. See, e.g., Ausubel et al., Current Protocols In Molecular Biology, Volumes 1 and 2 (John Wiley and Sons, New York 1989) and Sambrook et al., Molecular

Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, New York 1989). Testis cDNA libraries can also be purchased from ClonTech Laboratories, Inc., 1020 E. Meadow Circle, Palo Alto, CA 94303-4230.

The proteins of the invention can be used in contraceptive vaccines in mammals. Preferably a protein from the same species of mammal that is to be immunized is used in the vaccine. However, given the expected close homology of the proteins from different mammalian species, it is expected that proteins from other species, especially closely-related species, can be used.

Immunogenic portions of the proteins can also be us d in the vaccines. Immunogenic portions of the proteins must include at least a B-cell epitope. In choosing an immunogenic portion of testis-specific calpastatin, a portion must be chosen which includes sequences found on the testis-specific isoform but not found on the somatic isoforms.

Further, care should taken in using testis-specific calpastatin, or an immunogenic portion thereof, since somatic isoforms exist, and cross-reaction with these somatic isoforms may occur if the complete protein or an immunogenic portion containing an immunogenic somatic sequence is used in the vaccine. This may cause deleterious side effects and should be avoided except when the vaccine is to be used for contraception in pest species (e.g., rodents).

Preferably peptides derived from the proteins of the invention are used in the vaccines. To produce antibodies that react specifically with one of the proteins of the invention, the peptides must comprise at least a B-cell epitope of the protein. A peptide derived from testis-specific calpastatin must include a B-cell epitope from the sequences found on the testis-specific isoform but not found on the somatic isoforms. The peptide may include other sequences besides those which form the B-cell epitope, but these sequences must be chosen so that the

antibody produced as a result of immunization with the vaccin containing the p ptide will react specifically with the protein found in testis and sperm.

Methods of identifying B-cell epitopes of a protein are known. See O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993); O'Hern and Goldberg, Proceed. Intern. Symp. Control Rel. Bioact. Mater., 20, 394-395 (1993). Three criteria are essential for immunogenicity: a size greater than 10 amino acids; surface accessibility of the sequence; and hypervariability (degree of foreignness). See O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993); O'Hern and Goldberg, Proceed. Intern. Symp. Control Rel. Bioact. Mater., 20, 394-395 (1993).

The human testis-specific isoform of calpastatin has the following sequence at its N-terminal:

Met Gly Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser Pro
5 10

Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn Ala Glu
15 20 25

Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr Lys Gln
30 35 40

SEQ ID NO:1.

This sequence of 41 amino acids is unique to the testis-specific isoform of calpastatin. Peptides having this sequence, or a portion of it that includes the sequence from amino acid 26 through amino acid 41, can be used to elicit antibodies that react with the testis-specific isoform of calpastatin, but do not react with somatic isoforms of calpastatin. Amino acids 26-41 in the above sequence have been identified as a B-cell epitope.

The protein coded for by clone C-2 contains the following sequence:

Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu Asp Pro Thr Phe
15 20 25

Glu

SEQ ID NO:8.

Peptides having this sequence, or a portion of it that includes the sequence from amino acid 4 through amino acid 17, can be used to elicit antibodies that react specifically with the C-2 protein. Amino acids 4-17 in the above sequence have been identified as a B-cell epitope.

The protein coded for by clone L-7 contains the following sequence:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala
15 20

SEQ ID NO:11

and the following sequence:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn
15

SEQ ID NO:12.

Both of these sequences of amino acids (SEQ ID NO:11 and SEQ ID NO:12) have been identified as B-cell epitopes, and peptides having these sequences can be used to elicit antibodies that react specifically with the protein.

The peptides comprising a B-cell epitope of one of the proteins of the invention are preferably used in the vaccines in the form of an immunogen comprising the peptide

linked to a carrier. Suitable carriers are compounds capable of stimulating the production of antibodies to haptens coupled to them in a host animal. Many such carriers are well-known.

For instance, the carrier may be a high molecular weight compound. Suitable high molecular weight compounds include proteins, polypeptides, carbohydrates, polysaccharides, lipopolysaccharides, nucleic acids, and the like of sufficient size and immunogenicity.

Preferred high molecular weight compounds are proteins and polypeptides. Suitable immunogenic carrier proteins and polypeptides will generally have molecular weights between 4,000 and 10,000,000, and preferably greater than 15,000. Such suitable carriers include proteins such as albumins (e.g., bovine serum albumin, ovalbumin, human serum albumin), immunoglobulins, thyroglobulins (e.g., bovine thyroglobulin), hemocyanins (e.g., Keyhole Limpet hemocyanin), toxins (e.g., diptheria toxoid, tetanus toxoid) and polypeptides such as polylysine or polyalaninelysine. Preferred are diptheria toxoid and tetanus toxoid.

Methods of coupling the peptides to high molecular weight carriers are well-known. For instance, the peptide may be coupled to the carrier with conjugating reagents such as glutaraldehyde, a water soluble carbodiimide such as 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (ECDI), N-N-carbonyldiimidazole, 1--hydroxybenzotriazole monohydrate, N-hydroxysuccinimide, 6-maleimidocaproyl-N-hydroxysuccinimide, n-trifluoroacetylimidazole cyanogen bromide, 3-(2'--benzothiazolyl-dithio) propionate succinimide ester hydrazides or affinity labeling methods. See also Pierce Handbook and General Catalog (1989) for a list of possible coupling agents.

Additional references concerning conventional high molecular weight immunogenic carrier materials and techniques for coupling haptens thereto are: Erlanger,

Methods In Enzymology, 70, 85-104 (1980); Makela and Seppala, Handbook of Experimental Immunology (Blackwell 1986); Parker, Radioimmunoassay of Biologically Active Compounds (Prentice-Hall 1976); Butler J. Immunol. Meth., 7, 1-24 (1974); Weinryb and Shroff, Drug. Metab. Rev., 10, 271-83 (1979); Broughton and Strong, Clin. Chem., 22, 726-32 (1976); Playfair et al., Br. Med. Bull., 30, 24-31 (1974); U.S. Patents Nos. 4,990,596 and 4,782,136.

The number of peptides attached to the high molecular weight carrier is called the "epitopic density." The epitopic density can range from 1 to the number of available coupling groups on the carrier molecule. The epitopic density on a particular carrier will depend upon the molecular weight of the carrier and the density and availability of coupling sites. Preferably, only high molecular weight carriers having an epitopic density of at least 15 peptides per molecule are used in the vaccines of the invention.

The carrier may also be a peptide which has a sequence comprising the sequence of a T-cell epitope of one of the proteins of the invention or of another protein. Methods of identifying T-cell epitopes are known. See, O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993); O'Hern and Goldberg, Proceed. Intern. Symp. Control Rel. Bioact. Mater., 20, 394-395 (1993). The three criteria for selection of a T-cell epitope are: a size of 8-12 amino acids; hypervariability; and one or more representations of the tetrapeptide motif previously reported to be associated with T-cell epitopes. O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993); O'Hern and Goldberg, Proceed. Intern. Symp. Control Rel. Bioact. Mater., 20, 394-395 (1993).

Most preferably the carrier is a peptide which has a sequence comprising the sequence of a promiscuous T-cell epitope. A promiscuous T-cell epitope is a T-cell epitope that is recognized by individuals of several different major histocompatability (MHC) types. Promiscuous T-cell

epitopes are known. See, Ho et al., Eur. J. Immunol., 20, 477-483 (1990); Kaumaya, et al., J. Molec. Recog., 6, 81-94 (1993). A preferred promiscuous T-cell epitope has the following sequence:

Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

SEQ ID NO:5.

A peptide carrier which has a sequence comprising the sequence of a T-cell epitope may include other sequences linked to the N-terminal or C-terminal of the T-cell epitope. In particular, additional amino acids may be provided to link the B-cell epitope on the peptide to the T-cell epitope on the carrier. These linking amino acids should form a four-residue β-turn based on examination of 33 patterns in native proteins that code for $\alpha\alpha$ corners. Efimov, FEBS Lett., 166, 33 (1984); Kaumaya et al., Biochemistry, 29, 13-23 (1990).

Peptides comprising a B-cell epitope may be coupled to a peptide carrier comprising a T-cell epitope in the same manner as described above for high molecular weight proteins and polypeptides to form the immunogen. However, such immunogens are preferably synthesized as a single peptide in the ways described below for the synthesis of peptides.

The vaccines contain one or more of the proteins (or an immunogenic portion thereof), peptides and immunogens of the invention in a delivery system. Suitable delivery systems are well known. For instance, the delivery system may simply be a solvent (such as saline and buffers) or other liquid (such as an oil). However, the delivery system preferably enhances the immune response. Such delivery systems include aluminum salts, water-oil emulsions (such as incomplete Freund's adjuvant), saponins, liposomes, immune stimulating complex, lipopolysaccharides,

mycobacterial adjuvants (such as Freund's complete adjuvant), Squalen -Arlacel A containing the synthetic muramyl dipeptide N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP11637; Ciba-Geigy Pharmaceuticals, Basel, Switzerland), live vectors, antigen immunotargeting materials, and polymers (e.g., biodegradable microspheres, such as polylactide-polyglycolide microspheres, and block copolymers for sustained release). See Goldberg, in Gamete Interaction: Prospects For Immunocontraception, pages 63-73 (1990); Alexander et al., Reprod. Fertil. Dev., 6, 273-80 (1994); O'Hern et al., Biol. Reprod., 52, 331-339 (1995).

The vaccines may be administered in any conventional manner, including orally, intradermally, subcutaneously, intramuscularly, etc. to male or female mammals to inhibit fertilization of eggs by sperm. Suitable routes of administration and effective amounts (effective dosages and number of doses) necessary to inhibit conception can be determined empirically as is known in the art. By "inhibit" is meant at least a 50% reduction in the number of female mammals becoming pregnant as a result of the administration of the vaccine. Preferably at least a 75%, most preferably at least a 90%, reduction is achieved.

The proteins and peptides comprising a B-cell epitope can also be used in assays to assess infertility. The peptides may used as such or may be linked to a carrier. The carriers (e.g., large molecular weight and T-cell epitope carriers) and methods of linking the peptides to the carriers are the same as described above for the immunogens. To perform the assay, the protein, peptide or peptide linked to a carrier is contacted with a body fluid of a patient under conditions that permit antibodies in the body fluid to bind to it. Thus, the assays are immunoassays that allow for the determination of whether the body fluid of a patient contains antibodies that bind to the protein, peptide or peptide linked to a carrier. Suitable immunoassays and reagents for use therein are well

known in the art, and those skilled in the art will be able to determine operative and optimal assay conditions using only ordinary skill in the art.

Preferably the protein, peptide or peptide linked to a carrier will be immobilized on a solid surface. Suitable solid surfaces are well-known and include glass, polystyrene, polypropylene, polyethylene, nylon, paper, fiberglass, polyacrylamide and agaroses. The immobilized material is contacted with the body fluid so that antibodies present in the body fluid can bind to the protein, peptide or peptide linked to a carrier. After washing away unbound materials, a labeled secondary antibody or other material which binds specifically to the antibody in the body fluid is added as a means to detect and quantitate the antibody bound to the protein, peptide or peptide linked to a carrier. Suitable labels are well known in the art. They include enzymes, fluorophores, radionucleotides, bioluminescent labels, chemiluminescent labels, and particulate labels. The binding and detection of these labels can be accomplished using standard techniques well known to those skilled in the art.

The body fluid may be any body fluid that contains antibodies. Suitable body fluids include serum, plasma, cervical mucus and seminal plasma.

The assays may be used to assess infertility in patients unable to conceive. If the patient has antibodies specific for one of the proteins of the invention, then this may be the cause, or one of the causes, of the infertility. The assays may also be used to evaluate whether administration of the vaccines of the invention has been effective in immunizing recipients of the vaccines.

The invention also comprises a kit. The kit is a packaged combination of one or more containers holding reagents useful in performing the immunoassays. Suitable containers for the reagents include bottles, vials, test tubes, microtiter plates, a solid phase (see listing above) held in a molded plastic device, and other containers known

in the art. The kit will contain at least one container holding a prot in, peptide c mprising a B-cell epitope or such a peptide linked to a carrier. The kit may also comprise a container of a labeled component useful for detecting or quantitating the antibodies in the body fluids that bind to the protein, peptide or peptide linked to a carrier. The kit may also contain other materials which are known in the art and which may be desirable from a commercial and user standpoint, such as buffers, enzyme substrates, diluents, standards, etc. Finally, the kit may include containers, such as test tubes and microtiter plates, for performing the immunoassay.

The peptides of the invention may be made in a variety of ways. For instance, solid phase synthesis techniques may be used. Suitable techniques are well known in the art, and include those described in Merrifield, in *Chem. Polypeptides*, pp. 335-61 (Katsoyannis and Panayotis eds. 1973); Merrifield, *J. Am. Chem. Soc.*, 85, 2149 (1963); Davis et al., Biochem. Int'l, 10, 394-414 (1985); Stewart and Young, *Solid Phase Peptide Synthesis* (1969); U.S. Patents Nos. 3,941,763, 4,782,136, 4,990,596; Finn et al., in *The Proteins*, 3rd ed., vol. 2, pp. 105-253 (1976); and Erickson et al. in *The Proteins*, 3rd ed., vol. 2, pp. 257-527 (1976). Solid phase synthesis is the preferred method of making the peptides of the invention.

The peptides may also be produced by culturing a host cell comprising a DNA molecule coding for the peptide operatively linked to expression control sequences under conditions permitting expression of the peptide. The proteins of the invention may also be produced in this manner. In particular, the proteins and peptides can be produced in transformed host cells using recombinant DNA techniques. Such techniques and suitable host cells and other reagents for use therein are well known in the art.

For instance, the selection of a particular host cell is dependent upon a number of factors recognized by the art. These include, for example, compatibility with the

chosen expression vector, use and toxicity of the pr tein or peptide encoded by the expression vector, rate of transformation, expression characteristics, bio-safety, and costs. A balance of these factors must be struck with the understanding that not all hosts may be equally effective for the expression of a particular protein or peptide. Within the above guidelines, useful host cells include bacteria, yeast and other fungi, animal cell lines, animal cells in an intact animal, or other host cells known in the art.

The host cells may be transformed with a vector comprising DNA encoding the peptide or protein. On the vector, the coding sequence must be operatively linked to a promoter. The promoter used in the vector may be any sequence which shows transcriptional activity in the host cell and may be derived from genes encoding homologous or heterologous proteins and either extracellular or intracellular proteins, such as amylase, glycoamylases, proteases, lipases, cellulases, and glycolytic enzymes.

However, the promoter need not be identical to any naturally-occurring promoter. It may be composed of portions of various promoters or may be partially or totally synthetic. Guidance for the design of promoters is provided by studies of promoter structure such as that of Harley and Reynolds, Nucleic Acids Res., 15, 2343-61 (1987). Also, the location of the promoter relative to the transcription start may be optimized. See Roberts, et al., Proc. Natl Acad. Sci. USA, 76, 760-4 (1979).

The promoter may be inducible or constitutive, and is preferably a strong promoter. By "strong," it is meant that the promoter provides for a high rate of transcription in the host cell.

In the vector, the coding sequences must be operatively linked to transcription termination sequences, as well as to the promoter. The coding sequence may also be operatively linked to expression control sequences other than the promoters and transcription termination sequences.

These additional xpression control sequences include activators, enhancers, operators, stop signals, cap signals, polyadenylation signals, ribosome binding sites, and other signals involved with the control of transcription and translation.

In prokaryotic mRNA, the site at which the ribosome binds to the messenger includes a sequence of 3-9 purines. The consensus sequence of this stretch is 5'-AGGAGG-3', and it is frequently referred to as the Shine-Dalgarno sequence. The sequence of the ribosome binding site may be modified to alter expression. See Hui and DeBoer, Proc. Natl. Acad. Sci. USA, 84, 4762-66 (1987). Comparative studies of ribosomal binding sites, such as the study of Scherer, et al., Nucleic Acids Res., 8, 3895-3907 (1987), may provide guidance as to suitable base changes.

The ribosome binding site lies 3-12 bases upstream of the start (AUG) codon. The exact distance between the ribosome binding site and the translational start codon, and the base sequence of this "spacer" region, affect the efficiency of translation and may be optimized empirically.

To achieve optimal expression of a protein or peptide in prokaryotes, a ribosome binding site and spacer that provide for efficient translation in the prokaryotic host cell should be provided. A preferred ribosome binding site and spacer sequence for optimal translation in E. coli are described in Springer and Sligar, Proc. Nat'l Acad. Sci. USA, 84, 8961-65 (1987) and von Bodman et al., Proc. Nat'l Acad. Sci. USA, 83, 9443-47 (1986). The sequence of this ribosome binding site and spacer is: AGGAGAACAA CAACC [SEQ ID NO:28].

The consensus sequence for the translation start sequence of eukaryotes has been defined by Kozak (Cell, 44, 283-292 (1986)) to be: C(A/G)CCAUGG. Deviations from this sequence, particularly at the -3 position (A or G), have a large effect on translation of a particular mRNA. Virtually all highly expressed mammalian genes use this

sequence. Highly expressed yeast mRNAs, on the other hand, diff r from this sequence and instead use the s quence (A/Y)A(A/U)AAUGUCU (Cigan and Donahue, Gene, 59, 1-18 (1987)). These sequences may be altered empirically to determine the optimal sequence for use in a particular host cell.

Methods of preparing DNA molecules are well known in the art. For instances, sequences coding for the protein or peptide could be excised from genes or cDNA clones by methods well known in the art. However, the DNA molecules encoding a protein or peptide of the invention are preferably chemically synthesized. Methods of chemically synthesizing DNA are well known in the art. Chemical synthesis is preferable for several reasons.

First, chemical synthesis is desirable because codons preferred by the host in which the DNA sequence will be expressed may be used to optimize expression. Not all of the codons need to be altered to obtain improved expression, but greater than 50%, most preferably at least about 80%, of the codons should be changed to host-preferred codons. The codon preferences of many host cells, including E. coli, yeast, and other prokaryotes and eukaryotes, are known. See Maximizing Gene Expression, pages 225-85 (Reznikoff & Gold, eds., 1986). The codon preferences of other host cells can be deduced by methods known in the art.

The use of chemically synthesized DNA also allows for the selection of codons with a view to providing unique or nearly unique restriction sites at convenient points in the sequence. The use of these sites provides a convenient means of constructing the synthetic coding sequences. In addition, if secondary structures formed by the messenger RNA transcript interfere with transcription or translation, they may be eliminated by altering the codon selections.

Chemical synthesis also allows for the use of optimized expression control sequences with the DNA sequence coding for a protein or peptide. In this manner,

optimal expression of the prot in or peptide can be obtained. For instance, as noted above, promot rs can be chemically synthesized and their location relative to the transcription start optimized. Similarly an optimized ribosome binding site and spacer can be chemically synthesized and used with coding sequences that are to be expressed in prokaryotes.

DNA coding for a signal or signal-leader sequence may be located upstream of the DNA sequence encoding the protein or peptide. A signal or signal-leader sequence is an amino acid sequence at the amino terminus of a protein which allows the protein to which it is attached to be secreted from the cell in which it is produced. Suitable signal and signal-leader sequences are well known. Although secreted proteins are often easier to purify, secretion is generally not preferred since expression levels are much lower than those that can be obtained in the absence of secretion.

The vector used to transform the host cells may have one or more replication systems which allow it to replicate in the host cells. In particular, when the host is a yeast, the vector should contain the yeast 2u replication genes REP 1-3 and origin of replication. Many bacterial replicons are known.

Alternatively, an integrating vector may be used which allows the integration into the host cell's chromosome of the sequence coding for the protein or peptide. Although the copy number of the coding sequence in the host cells would be lower than when self-replicating vectors are used, transformants having sequences integrated into their chromosomes are generally quite stable.

When the vector is a self-replicating vector, it is preferably a high copy number plasmid so that high levels of expression are obtained. As used herein, a "high copy number plasmid" is one which is present at about 100 copies or more per cell. Many suitable high copy number plasmids are known.

The vector desirably also has unique restriction sites for the insertion of DNA sequences and a sequence coding for a selectable or identifiable phenotypic trait which is manifested when the vector is present in the host cell ("a selection marker"). If a vector does not have unique restriction sites, it may be modified to introduce or eliminate restriction sites to make it more suitable for further manipulations.

After the vector comprising the sequence coding for the protein or peptide is prepared, it is used to transform the host cells. Methods of transforming host cells ar well known in the art, and any of these methods may b used. Transformed host cells are selected in known ways and then cultured to produce the protein or peptide.

The methods of culture are those well known in the art for the chosen host cell, but the use of enriched media (rather than minimal media) is preferred since higher yields are obtained. The expressed protein or peptide may be recovered using methods of recovering and purifying proteins from cell cultures which are well known in the art.

EXAMPLES

EXAMPLE 1: Identification Of Testis-Specific Clones

A human testis cDNA library was screened with sera from infertile patients positive for antisperm antibodies. This screening was performed as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994). It is interesting to note that these patients, although infertile, were otherwise healthy.

A total of 43 unique cDNA inserts were detected by the screening, of which four were testis-specific by Northern blot analysis (performed as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994); see below). One of the four clones turned out to encode a truncated mRNA for a somatic peptide and was not evaluated further. The remaining three clones were designated Y-19, C-2 and L-7.

EXAMPLE 2: Characterization Of Clone Y-19

1. DNA Sequence

The sequence of the cDNA insert of clone Y-19 was determined as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994). The DNA sequence of the insert and the deduced corresponding amino acid sequence are set forth in Chart A below.

Homology searches of the GenEMBL databases (performed as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994)) indicated that clone Y-19 codes for a testis-specific isoform of human calpastatin.

Figure 1 shows the relationship between the published sequence of DNA coding for somatic calpastatin (solid) and the testis-specific region of clone Y-19 (diagonal stripes). Clone Y-19 appears to be a product of alternative splicing whereby DNA coding for somatic calpastatin domains L and 1 has been deleted and replaced with DNA coding for a unique, testis-specific L domain of approximately 65 amino acids (stripes). The rest of the cDNA sequence of clone Y-19 is virtually identical to the

published sequence of somatic calpastatin. However, DNA coding for testis-specific calpastatin contains 2 unique restriction sites (arrows).

2. Northern Blots

Northern blots were performed as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994).

A 1kb fragment of clone Y-19 was used to probe a Northern blot of human poly A+ RNA from eight different human tissues (leukocytes, colon, small intestine, ovary, testis, prostate, thymus and spleen; Multiple Tissue Northern blots purchased from Clonetech, Palo Alto, CA). Two mRNAs of 4.3 and 2.8kb were detected by the probe in all tissues. A third mRNA of 1.9kb was detected only in testis.

The Multiple Tissue Northern blots probed with the 1kb Y-19 fragment were stripped as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994) and re-probed with a 135 bp fragment of the unique 5' sequence of Y-19. Only the 1.9kb mRNA in testis was detected with this probe.

3. Serum YM

The serum that identified clone Y-19 (serum YM) agglutinates human sperm in a head-to-head orientation and completely inhibits cervical mucus penetration.
These assays were performed as described in Schulman et al., Am. J. Obstet Gynecol., 123, 139-144 (1975) and Ansbacher et al., Fertil. Steril., 24, 305-308 (1973).

EXAMPLE 3: Identification Of B-Cell Epitope Of Testis-Specific Calpastatin

The complete amino acid sequence of human testis-specific calpastatin coded for by clone Y-19 is set forth in Chart A below. A comparison of the first 41 amino acids of human somatic calpastatin with the first 41 residues of human testis-specific calpastatin showed no sequence homology between them:

SEQ ID NO:15

Somatic: MNPTETKAIPVSQQMEGPHLPNKKKHKKQAVKTEPEKKSQS

Testis-
Specific: MGQFLSSTFLEGSPATVSTISFVTVNAEEQEKQFVSSRTKQ

SEQ ID NO:1

Beginning at residue 42 of testis-specific calpastatin (residue 387 of somatic calpastatin), the two sequences are virtually identical.

Figure 2 shows a computer-generated hydropathy plot of the first 41 residues of somatic calpastatin (solid lines) versus the first 41 residues of testis-specific calpastatin (open bars). This hydropathy plot was generated using algorithms described in Hopp and Woods, Proc. Natl. Acad. Sci. USA, 78, 3824-28 (1981) and Kyte and Doolittle, J. Mol. Biol., 157, 105 (1982). Only residues 26-41 of testis-specific calpastatin are both hydrophilic and unique to the testis isoform. Therefore, this segment was chosen as a testis-specific B-cell epitope. This segment has the sequence:

Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr
5 10

Lys Gln
15

SEQ ID NO:2.

The hydropathy plot also shows that testis-specific calpastatin has a hydrophobic tail. This hydrophobic tail could serve as a membrane anchor for the protein.

EXAMPLE 4: Preparation Of Immunogen Containing B-Cell Epitope Of Testis-Specific Calpastatin And Uses Thereof

A peptide immunogen was prepared containing the testis-specific calpastatin B-cell epitope identified in Example 3 linked to a carrier comprising a universal T-cell epitope derived from tetanus toxoid. The T-cell epitope had the following sequence:

Val Asp Asp Ala Leu Il Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

SEQ ID NO:5.

Four amino acids (Gly Pro Ser Leu) were used to link the B-cell epitope to the T-cell epitope. Thus, the complete carrier sequence was:

Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser Thr Lys
5 10

Ile Tyr Ser Tyr Phe Pro Ser Val
15 20

SEQ ID NO:6,

and the complete immunogen had the following sequence:

Thr Val Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser
5 10

Arg Thr Lys Gln Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

SEQ ID NO:7.

This immunogen [SEQ ID NO:7] was synthesized at the Salk Institute (under Contract N01-HD-0-2906 with the NIH) and made available by the Contraceptive Development Branch, Center for Population Research, NICHD (Bethesda, MD).

Female New Zealand White rabbits were immunized with the immunogen [SEQ ID NO:7] as described in O'Hern et al., Biol. Reprod., 52, 331-339 (1995). The rabbit antiserum was affinity purified by epitope selection as described in Snyder et al., Methods Enzymol., 154, 107-128 (1987).

The affinity-purified antiserum was used to probe a Western blot of human tissue extracts. The tissue extracts were made and the Western blots were p rformed as d scribed

in Diekman and Goldberg, Biol. Reprod., 50, 1087-1093 (1994). As shown in Figure 3, the antiserum recognized a single protein of approximately 65Kd in human testis extracts (lane 1) and a slightly larger protein of approximately 68Kd in human sperm extracts (lane 2). There was no reactivity with human liver extracts (lane 3), although liver is known to be rich in the somatic isoforms of calpastatin.

The affinity-purified antiserum was also used to localize testis-specific calpastatin on human sperm by immunofluorescence, performed as described in Wright et al., Biol. Reprod., 42, 693-701 (1990). Diffuse, granular fluorescence was observed throughout the acrosome, and intense fluorescence was observed in the equatorial segment of the sperm.

EXAMPLE 5: Immunization With Immunogen Containing B-Cell Epitope Of Testis-Specific Calpastatin

Female cynomologous macaques (three per group) were immunized with either 100μg or 300μg of the peptide immunogen [SEQ ID NO:7] prepared in Example 4. The immunogen was administered intramuscularly in Squalene-Arlacel A containing the synthetic muramyl dipeptide N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP11637; Ciba-Geigy Pharmaceuticals, Basel, Switzerland). A single booster injection consisting of the same dose in the same delivery system was administered intramuscularly ten days after the initial injection.

ELISA titers were determined on microtiter plates coated with the testis-specific calpastatin B-cell epitope peptide (SEQ ID NO:2; see Example 3) conjugated to bovine serum albumin (BSA). The B-cell epitope peptide was synthesized with a non-natural cysteine at the amino terminus and conjugated to BSA as described in O'Hern et al., Biol. Reprod., 52, 331-339 (1995). The ELISA was performed as described in Laerimore et al., J. Virol., 69, 6077-6089 (1995). The microtiter plate was coated with

peptide-conjugated BSA or BSA alone. After standard washing and blocking procedures, goat anti-human IgG conjugated to horseradish peroxidase was added to detect bound antibody. The results were recorded as absorbance of duplicate wells minus background absorbance. The results are shown in Figure 4 where open symbols denote the low dose group (100μg), closed symbols denote the high dose group (300 μg), and the arrows show the time of the booster injections.

EXAMPLE 6: Characterization Of Clone C-2

The cDNA insert of clone C-2 was used to probe a Northern blot of human poly A+ RNA from eight different human tissues as described above in Example 2. A single mRNA of 2.1kb was detected in testis only.

The sequence of the cDNA insert of clone C-2 was determined as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994). The DNA sequence of the insert and the deduced corresponding amino acid sequence are set forth in Chart B below.

Homology searches of the GenEMBL databases found that the sequence of the cDNA insert of clone C-2 was not represented. Thus, clone C-2 cDNA encodes a unique and previously undescribed protein.

As noted above, the mRNA is approximately 2.1 kb. It has an open reading frame (ORF) of 1.4 kb translating to a peptide of 65-70 Kd. There are no significant sequence motifs or unusual properties.

The original antiserum that detected clone C-2 (number 629) is 100% effective in blocking fertilization *in vitro* of human ova by human sperm (see table below). Serum 629 which has been absorbed with sperm no longer blocks binding of sperm to zona (see table below). These assays were performed by Gary Clarke, The Royal Womens' Hospital, Melbourne, Australia, using procedures described in Clarke et al., Arch. Androl., 35, 21-27 (1995).

Serum Treatment	Number Ova Fertilized	Number Sperm Bound To Zona
Normal Serum	5/6	62
629	0/10	1.5
629 Preabsorbed With Sperm	ND	67

The peptide coded for by a 900 bp fragment from the 3' end of the C-2 cDNA was expressed as a glutathione-s-transferase (GST) fusion protein using cloning methods well known in the art. See, e.g., Smith and Johnson, Gene, 67, 31-40 (1988); Johnson et al., Nature, 338, 585-587 (1989); Kemp et al., Gene, 94, 223-28 (1990); Kaelin Jr. et al., Cell, 64, 521-532 (1991); Chittenden Jr. et al., Cell, 65, 1073-1082 (1991); Kaelin Jr. et al., Cell, 70, 351-364 (1992). The clone encoding this fusion protein was designated clone GST-C2.

Western blots (performed as described above in Example 4) showed that the fusion protein was recognized by the 629 serum. It was not recognized by the 629 serum which had been absorbed with human sperm. Furthermore, the sera from four other infertile patients recognized this fusion protein on Western blots. One of these sera inhibited sperm-zona binding.

EXAMPLE 7: Identification Of B-Cell Epitope Of Clone C-2 Protein

Unidirectional nested deletions were prepared from the 3' end of clone GST-C2 (see Figure 5, upper portion) using the protocol and reagents provided in the Stratagene instruction manual (pBluescript II exo/mung DNA sequencing system). Each time point was religated, and the truncated GST-C2 fusion proteins were expressed and assayed by PAGE as described in the previous example. The lower half of Figure 5 shows the Coomasie blue-stained PAGE gel (lanes 1 and 7 - GST, lane 2 - full-length GST-C2 fusion protein, lanes 3-6 and 8-11 - truncated GST-C2 fusion proteins).

Each of the truncated GST-C2 fusion proteins was partially purified and used as the target for Western blots (all as described in Example 6) probed with the original patient 629 serum. The results are shown in Figure 6. The full-length fusion protein and the first 4 deletions were strongly positive for the antibody. Time points 5-10 were negative, as was GST alone. Therefore, the C2 epitope recognized by the original human serum resides within time point 4.

Each of the 10 nested deletions was sequenced using an oligo primer specific for the pGEX vector (see Pharmacia Biotech GST Gene Fusion Manual). The results are shown in Figure 7. The first 3 time points showed deletion of the 3' untranslated region (UTR). Time point 4, from which the 9 carboxy terminal amino acids were deleted, was still antibody positive. Time point 5, with deletion of an additional 26 amino acids, was antibody negative. Therefore, the relevant B-cell epitope (cross-hatched box) resides within the region of amino acids 426-454. The sequence of amino acids 426-454 is as follows:

Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu Asp Pro Thr Phe
15 20 25

Glu

SEQ ID NO:8

Computer-assisted sequence analysis was performed as described in O'Hern and Goldberg, in Techniques In Protein Chemistry IV, pages 481-490 (1993) to calculate the surface accessibility of amino acids 426-454. Residues 430-443 were determined to be highly surface accessible and likely to represent the B-cell epitope. This epitope has the following sequence:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys
15

SEQ ID NO:9.

EXAMPLE 8: Preparation of C-2 Immunogen

An immunogen comprising the B-cell epitopes identified in Example 7 was prepared as described in Example 4. The sequence of this immunogen is:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg Pro Glu
5 10

Pro Lys Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35

SEQ ID NO:10.

EXAMPLE 9: Characterization Of Clone L-7

The cDNA insert of clone L-7 was used to probe a Northern blot of human poly A+ RNA from eight different human tissues as described above in Example 2. A single mRNA of 2.5kb was detected in testis only.

The sequence of the cDNA insert of clone L-7 was determined as described in Liang et al., Reprod. Fertil. Dev., 6, 297-305 (1994). The DNA sequence of the insert and the corresponding amino acid sequence are set forth in Chart C below.

Homology searches of the GenEMBL databases found that the sequence of the cDNA insert of clone L-7 was not represented. Thus, clone L-7 cDNA encodes an unique and previously undescribed protein. This protein is relatively large (66 kD) and consists of several domains of as yet unknown functional significance. The protein contains an endoplasmic reticulum signal sequence and appears to be

anchored in the sperm plasma membrane at its amino terminus, but with surface accessible epitopes.

A computer-generated plot (Figure 8) of the occurrence of the amino acid valine along the length of the polypeptide chain revealed a distinct domain structure for the protein. This plot was generated using PC/Gene software from Intelligenetics, Inc., 700 E. El Camino Rd., Mountainview, CA 94047. This computer analysis revealed the following features. Residues 88-328 contain very little valine and 9 potential protein kinase C (PKC) phosphorylation sites (P). Residues 329 to 493 contains many valines and no PKC phosphorylation sites. Residues 329-493 also contain 11 repeats of a 15 amino acid motif (see below). The consensus sequence of the motif is KgqEaQVKKsesgVp [SEQ ID NO:16].

329-	KRTGVQVKKSESGVP	SEQ ID NO:17
344-	KGQEAQVTKSGLVVL	SEQ ID NO:18
359-	KGQEAQVEKSEMGVP	SEQ ID NO:19
374-	RRQESQVKKSQSGVS	SEQ ID NO:20
389-	KGQEAQVKKRESVVL	SEQ ID NO:21
404-	KGQEAQVEKSELKVP	SEQ ID NO:22
419-	KGQEGQVEKTEAECP	SEQ ID NO:23
434-	KEQEVQEKKSEAGVL	SEQ ID NO:24
449-	KGPEFQVKNTEVSVP	SEQ ID NO:25
464-	ETLESQVKKSESGVL	SEQ ID NO:26
479-	KGQEAQEKKESFEDK	SEQ ID NO:27

Residues 494-568 contain few valines and 3 potential PKC phosphorylation sites.

From the computer analysis and the protein's sequence, the following domain organization of the L-7 protein is proposed:

Domain I (residues 1-90) contains a consensus endoplasmic reticulum localization signal ($p>0.85$) (see von Heijne, J. Memb. Biol., 115, 195-201 (1990));

Domain II (residues 91-328) has a high isoelectric point and contains the 9 potential PKC phosphorylation sites;

Domain III (residues 329 to 493) has a neutral pI and contains the 11 repeat motifs; and

Domain IV (residues 494 to 568) again has a high isoelectric point and contains 2 bipartite nuclear translocation signals (see Robbins et al., Cell, 64, 615-623 (1991)).

This structure is unique in the databases.

EXAMPLE 10: Identification Of B-Cell Epitope Of Clone L-7 Protein

A 900 bp fragment from the 3' end of the cDNA of clone L-7 was expressed and purified as a GST fusion protein as described in Example 6 above. This clone was designated GST-L7. Sera from three infertile patients (numbers 44, 65 and 66) recognized the fusion protein on Western blots (performed as described in Example 6).

Nested deletions of the 900 bp fragment were prepared, and the truncated fusion proteins were expressed and purified, all as described in Example 7. Western blots were probed with serum from patient 44. The results are shown in Figure 9. Signal intensity decreased markedly between time points 2 and 3 (arrows) and disappears between time points 8 and 9 (arrows), indicating the presence of two B-cell epitopes in this region of the L-7 protein.

The two epitopes identified by nested deletion analysis of clone L-7 are indicated by cross-hatched boxes in Figure 10. Epitope 1 is amino acids 500-517, and epitope 2 is amino acids 389-408. These epitopes have the following sequences:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala
15 20

SEQ ID NO:11

and

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn
15 SEQ ID NO:12.

EXAMPLE 11: Preparation of L-7 Immunogens

Immunogens comprising the two B-cell epitopes identified in Example 10 were prepared as described in Example 4. The sequences of these two immunogens are:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala Gly Pro Ser Leu Val Asp Asp Ala
15 20 25

Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

SEQ ID NO:13.

and

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

SEQ ID NO:14.

EXAMPLE 12: Preparation Of Antiserum To L-7 Protein

One of the immunogens prepared in Example 11 [SEQ ID NO:14] was used to immunize rabbits as described in Example 4. The rabbit antiserum was affinity purified, and the affinity-purified rabbit antiserum was used to probe a

Western blot of human tissue extracts, all as described in Example 4. The affinity-purified antiserum recognized a single protein of approximately 58 Kd in human testis extracts and a protein of approximately 68 Kd in human sperm extracts. There was no reactivity with human liver extracts.

EXAMPLE 13: Isolation Of Macaque cDNA Clones Corresponding To Human cDNA Clones And Identification Of B-Cell Epitopes

A macaque testis cDNA library (obtained from Dr. John Herr, University of Virginia) was screened with the human cDNAs as probes (see Examples 1 and 2), and B-cell epitopes identified by comparison to B-cell epitopes identified in Examples 7 and 10.

A B-cell epitope of macaque testis-specific calpastatin was identified and has the following sequence:

Asn Ala Glu Gly Gln Glu Lys Gln Phe Leu Ser Ser Arg Thr
5 10

Lys Gln
15

SEQ ID NO:29.

This B-cell epitope is 85% homologous to the B-cell epitope identified above for human testis-specific calpastatin [SEQ ID NO:2].

The B-cell epitope of the macaque protein corresponding to the human protein produced by clone C-2 has a sequence identical to that of the B-cell epitope of the C-2 protein [SEQ ID NO:8]. Thus, in this case, there was 100% homology between the sequences.

EXAMPLE 15: Preparation Of Immunogens Containing Testis-Specific B-Cell Epitopes

Peptides having the sequences of the B-cell epitopes identified in Examples 3, 7 and 10 can be synthesized and coupled to diptheria toxin to produce immunogens that can

b used to immunize mammals, all as described in O'Hern et al., Biol. Reprod., 52, 331-339 (1995).

EXAMPLE 16: Sequencing Of Clones Y-19, C-2 and L-7

DNA fragments of clones Y-19, C-2 and L-7 were subcloned into the pBluescriptII SK+ phagemid (Stratagene, Palo Alto, CA) and sequenced by a modification of the method of Kraft et al., Biotechniques, 6, 544-547 (1988) as described in O'Hern et al., Biol. Reprod., 52, 331-339 (1995). The DNA sequences and deduced amino acid sequences are presented in Charts A (Y-19), B (C-2) and C (L-7).

CHART A

```
CTTGATATCG AATTCGGGGGG AGTCTCCCT GACTTCCAGC                    40

AACAATCCTT GAGTCTGAGA CTGCCCTGGC CTAAG ATG GGC                 81
                                        Met Gly

CAG TTT CTA TCT TCG ACT TTC TTG GAG GGC TCA CCG                117
Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser Pro
          5                  10

GCC ACA GTG TCG ACG ATA AGC TTT GTG ACG GTG AAC                153
Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn
15                  20                  25

GCA GAG GAG CAA GAG AAG CAG TTC GTA TCT TCC AGG                189
Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg
              30                  35

ACC AAG CAA AAA GCT AAA GAA GAA AAA CTA GAG AAG                225
Thr Lys Gln Lys Ala Lys Glu Glu Lys Leu Glu Lys
    40                  45                  50

TGT GGT GAG GAT GAT GAA ACA ATC CCA TCT GAG TAC                261
Cys Gly Glu Asp Asp Glu Thr Ile Pro Ser Glu Tyr
                55                  60

AGA TTA AAA CCA GCC ACG GAT AAA GAT GGA AAA CCA                297
Arg Leu Lys Pro Ala Thr Asp Lys Asp Gly Lys Pro
          65                  70

CTA TTG CCA GAG CCT GAA GAA AAA CCC AAG CCT CGG                333
Leu Leu Pro Glu Pro Glu Glu Lys Pro Lys Pro Arg
75                  80                  85

AGT GAA TCA GAA CTC ATT GAT GAA CTT TCA GAA GAT                369
Ser Glu Ser Glu Leu Ile Asp Glu Leu Ser Glu Asp
              90                  95

TTC GAC CTG TCT GAA TGT AAA GAG AAA CCA TCT AAG                405
Phe Asp Leu Ser Glu Cys Lys Glu Lys Pro Ser Lys
    100                 105                 110

CCA ACT GAA AAG ACA GAA GAA TCT AAG GCC GCT GCT                441
Pro Thr Glu Lys Thr Glu Glu Ser Lys Ala Ala Ala
                115                 120

CCA GCT CCT GTG TCG GAG GCT GTG TCT CGG ACC TCC                477
Pro Ala Pro Val Ser Glu Ala Val Ser Arg Thr Ser
          125                 130

ATG TGT AGT ATA CAG TCA GCA CCC CCT GAG CCG GCT                513
Met Cys Ser Ile Gln Ser Ala Pro Pro Glu Pro Ala
135                 140                 145
```

```
ACC TTG AAG GTC ACA GTG CCA GAT GAT GCT GTA GAA   549
Thr Leu Lys Val Thr Val Pro Asp Asp Ala Val Glu
            150                 155

GCC TTG GCT GAT AGC CTG GGG AAA AAG GAA GCA GAT   585
Ala Leu Ala Asp Ser Leu Gly Lys Lys Glu Ala Asp
    160                 165                 170

CCA GAA GAT GGA AAA CCT GTG ATG GAT AAA GCT AAG   621
Pro Glu Asp Gly Lys Pro Val Met Asp Lys Val Lys
                175                 180

GAG AAG GCC AAA GAA GAA GAC CGT GAA AAG CTT GGT   657
Glu Lys Ala Lys Glu Glu Asp Arg Glu Lys Leu Gly
        185                 190

GAA AAA GAA GAA ACA ATT CCT CCT GAT TAT ATA TTA   693
Glu Lys Glu Glu Thr Ile Pro Pro Asp Tyr Ile Leu
195                 200                 205

GAA GAG GTC AAG GAT AAA GAT GGA AAG CCA CTC CTG   729
Glu Glu Val Lys Asp Lys Asp Gly Lys Pro Leu Leu
            210                 215

CCA AAA GAG TCT AAG GAA CAG CTT CCA CCC ATG AGT   765
Pro Lys Glu Ser Lys Glu Gln Leu Pro Pro Met Ser
    220                 225                 230

GAA GAC TTC CTT CTG GAT GCT TTG TCT GAG GAC TTC   801
Glu Asp Phe Leu Leu Asp Ala Leu Ser Glu Asp Phe
                235                 240

TCT GGT CCA CAA AAT GCT TCA TCT CTT AAA TTT GAA   837
Ser Gly Pro Gln Asn Ala Ser Ser Leu Lys Phe Glu
        240                 245

GAT GCT AAA CTT GCT GCT GCC ATC TCT GAA GTG GTT   873
Asp Ala Lys Leu Ala Ala Ala Ile Ser Glu Val Val
250                 255                 260

TCC CAA ACC CCA GCT TCA ACG ACC CAA GCT GGA GCC   909
Ser Gln Thr Pro Ala Ser Thr Thr Gln Ala Gly Ala
            265                 270

CCA CCC CGT GAT ACC TCG AGT GAC AAA GAC CTC GAT   945
Pro Pro Arg Asp Thr Ser Ser Asp Lys Asp Leu Asp
    275                 280                 285

GAT GCC TTG GAT AAA CTC TCT GAC AGT CTA GGA CAA   981
Asp Ala Leu Asp Lys Leu Ser Asp Ser Leu Gly Gln
                290                 300

AGG CAG CCT GAC CCA GAT GAG AAC AAA CCA ATG GAA  1017
Arg Gln Pro Asp Pro Asp Glu Asn Lys Pro Met Glu
        305                 310
```

```
GAT AAA GTA AAG GAA AAA GCT AAA GCT GAA CAT AGA                    1053
Asp Lys Val Lys Glu Lys Ala Lys Ala Glu His Arg
315                 320                 325

GAC AAG CTT GGA GAG AGA GAT GAC ACT ATC CCA CCT                    1089
Asp Lys Leu Gly Glu Arg Asp Asp Thr Ile Pro Pro
            330                 335

GAA TAC AGA CAT CTC CTG GAT GAT AAT GGA CAG GAC                    1125
Glu Tyr Arg His Leu Leu Asp Asp Asn Gly Gln Asp
    340                 345                 350

AAA CCA GTG AAG CCA CCT ACA AAG AAA TCA GAG GAT                    1161
Lys Pro Val Lys Pro Pro Thr Lys Lys Ser Glu Asp
                355                 360

TCA AAG AAA CCT GCA GAT GAC CAA GAC CCC ATT GAT                    1197
Ser Lys Lys Pro Ala Asp Asp Gln Asp Pro Ile Asp
        365                 370

GCT CTC TCA GGA GAT CTG GAC AGC TGT CCC TCC ACT                    1233
Ala Leu Ser Gly Asp Leu Asp Ser Cys Pro Ser Thr
375                 380                 385

ACA GAA ACC TCA CAG AAC ACA GCA AAG GAT AAG TGC                    1269
Thr Glu Thr Ser Gln Asn Thr Ala Lys Asp Lys Cys
            390                 395

AAG AAG GCT GCT TCC AGC TCC AAA GCA CCT AAG AAT                    1305
Lys Lys Ala Ala Ser Ser Ser Lys Ala Pro Lys Asn
    400                 405                 410

GGA GGT AAA GCG AAG GAT TCA GCA AAG ACA ACA GAG                    1341
Gly Gly Lys Ala Lys Asp Ser Ala Lys Thr Thr Glu
                415                 420

GAA ACT TCC AAG CCA AAA GAT GAC TAA AGAAATACAAG                    1377
Glu Thr Ser Lys Pro Lys Asp Asp
        425                 430

TTAAGGTATC TGGTATCTGC ATTTAAAATC TTCAGCTGGT                        1417

GGATTGTGAC TTTTGAAGAA CAAAAGGCTT TGGCAACAGA                        1457

AAACAATTGT TCTGGGTGAT TTCTAGAATG TTTTTTGTTG                        1497

AGTCTCTGAA CATCCTAAAT ATTTGTTTGT TATTCTTTTC                        1537

CAGAAAGAAA ATGAATTTGA CTGGTTCACC TGTGTACTGA                        1577

GTATTGATAA ACTTCGAATT TTTTAAATTT CCTTCAAGGG                        1617

AGAGAAAGCT TATATTGGTT TGTTATTCTT TTCCAGAAAG                        1657

AAAATGAATT TGACTGGGTT CACTGTGTTA CTGAGTATTG                        1697
```

ATAAACTTTG AATTTTTGCA ATTGCCTTCA ATTTTTAGAG 1737

GAAAAGCTTT ATATTTGTGT TATTACTTCT TCATCTTACA 1777

GTCATCACAG AACACACTGA GACTTGAATC AAGTCAGCAA 1817

CAGAGCAAAA TAAAGGTTAG ATAAGTCCTT GTGTAGCAAA 1857

TTTCGAGCAT AAGAAATAAA ATCTAATTAA TTCTTAGGGT 1897

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1927

SEQ ID NO:30

CHART B

```
AAAGCGTCAT TCGAGGTCCG GGTCCGGCTT GCGGGGTCAG                          40

CGAACTGGAG AGGCGCC  ATG GGC TGG ATC ACA                              72
                    Met Gly Trp Ile Thr
                                     5

GAA GAT CTT ATT AGA CGG AAT GCT GAA CAC AAC GAC                      108
Glu Asp Leu Ile Arg Arg Asn Ala Glu His Asn Asp
                10                  15

TGT GTC ATT TTT TCC CTG GAG GAA CTC TCG TTG CAT                      144
Cys Val Ile Phe Ser Leu Glu Glu Leu Ser Leu His
        20                  25

CAG CAA GAA ATA GAA AGA CTA GAA CAC ATT GAT AAA                      180
Gln Gln Glu Ile Glu Arg Leu Glu His Ile Asp Lys
30                  35                  40

TGG TGC CGG GAT TTA AAA ATT CTC TAT CTT CAA AAT                      216
Trp Cys Arg Asp Leu Lys Ile Leu Tyr Leu Gln Asn
            45                  50

AAT CTT ATT GGG AAA ATT GAA AAT GTT AGC AAA CTC                      252
Asn Leu Ile Gly Lys Ile Glu Asn Val Ser Lys Leu
    55                  60                  65

AAG AAA CTT GAA TAT TTG AAT TTA GCT TTA AAC AAC                      288
Lys Lys Leu Glu Tyr Leu Asn Leu Ala Leu Asn Asn
                70                  75

ATT GAA AAA ATA GAA AAC TTG GAA GGA TGT GAA GAG                      324
Ile Glu Lys Ile Glu Asn Leu Glu Gly Cys Glu Glu
        80                  85

CTG GCA AAA CTT GAC CTG ACT GTG AAT TTC ATT GGA                      360
Leu Ala Lys Leu Asp Leu Thr Val Asn Phe Ile Gly
90                  95                  100

GAG CTG AGC AGC ATT AAA AAC TTG CAG CAC AAT ATC                      396
Glu Leu Ser Ser Ile Lys Asn Leu Gln His Asn Ile
            105                 110

CAT CTG AAG GAG CTC TTT CTC ATG GGG AAC CCA TGT                      432
His Leu Lys Glu Leu Phe Leu Met Gly Asn Pro Cys
    115                 120                 125

GCT TCC TTT GAC CAC TAT AGG GAG TTC GTG GTA GCA                      468
Ala Ser Phe Asp His Tyr Arg Glu Phe Val Val Ala
                130                 135

ACT CTT CCA CAA TTA AAG TGG TTG GAT GGT AAA GAA                      504
Thr Leu Pro Gln Leu Lys Trp Leu Asp Gly Lys Glu
        140                 145
```

```
ATA GAG CCT TCA GAA AGG ATT AAG GCA TTG CAG GAC   540
Ile Glu Pro Ser Glu Arg Ile Lys Ala Leu Gln Asp
150                 155                 160

TAT TCA GTA ATT GAA CCA CAA ATC AGA GAG CAG GAA   576
Tyr Ser Val Ile Glu Pro Gln Ile Arg Glu Gln Glu
            165                 170

AAA GAT CAC TGT CTT AAA CGA GCC AAA CTC AAG GAA   612
Lys Asp His Cys Leu Lys Arg Ala Lys Leu Lys Glu
    175                 180                 185

GAG GCT CAG AGG AAA CAC CAA GAA GAG GAT AAA AAT   648
Glu Ala Gln Arg Lys His Gln Glu Glu Asp Lys Asn
                190                 195

GAA GAC AAG AGA AGT AAC GCA GGC TTT GAT GGA CGT   684
Glu Asp Lys Arg Ser Asn Ala Gly Phe Asp Gly Arg
        200                 205

TGG TAC ACA GAC ATC AAT GCT ACT CTT TCC TCT TTA   720
Trp Tyr Thr Asp Ile Asn Ala Thr Leu Ser Ser Leu
210                 215                 220

GAG AGC AAA GAC CAC CTA CAG GCA CCA GAC ATA GAG   756
Glu Ser Lys Asp His Leu Gln Ala Pro Asp Ile Glu
            225                 230

GAA CAC AAC ACA AAG AAA TTA GAC GAT GAC TTG GAA   792
Glu His Asn Thr Lys Lys Leu Asp Asp Asp Leu Glu
    235                 240                 245

TTC TGG AAT AAG CCC TGT TTG TTT ACT CCT GAA TCA   828
Phe Trp Asn Lys Pro Cys Leu Phe Thr Pro Glu Ser
                250                 255

AGA TTG GAA ACT CTT AGA CAC ATG GAA AAA CAA CGG   864
Arg Leu Glu Thr Leu Arg His Met Glu Lys Gln Arg
        260                 265

AAG AAA CAG GAA AAA TTA AGT GAA AAA AAG AAG AAA   900
Lys Lys Gln Glu Lys Leu Ser Glu Lys Lys Lys Lys
270                 275                 280

GTG AAA CCA CCC AGG ACT TTG ATC ACT GAA GAT GGG   936
Val Lys Pro Pro Arg Thr Leu Ile Thr Glu Asp Gly
            285                 290

AAA GCC CTA AAT GTG AAT GAG CCC AAA ATT GAC TTC   972
Lys Ala Leu Asn Val Asn Glu Pro Lys Ile Asp Phe
    295                 300                 305

TCT TTG AAA GAT AAC GAA AAG CAG ATC ATC CTG GAC  1008
Ser Leu Lys Asp Asn Glu Lys Gln Ile Ile Leu Asp
                310                 315
```

```
CTT GCT GTC TAT AGG TAT ATG GAT ACC TCT TTA ATC   1044
Leu Ala Val Tyr Arg Tyr Met Asp Thr Ser Leu Ile
        320                 325

GAT GTT GAT GTG CAA CCA ACT TAC GTG CGA GTA ATG   1080
Asp Val Asp Val Gln Pro Thr Tyr Val Arg Val Met
330                 335                 340

ATC AAA GGA AAG CCA TTT CAG CTT GTC CTT CCT GCA   1116
Ile Lys Gly Lys Pro Phe Gln Leu Val Leu Pro Ala
            345                 350

GAA GTG AAA CCC GAT AGT AGT TCT GCT AAA AGA TCT   1152
Glu Val Lys Pro Asp Ser Ser Ser Ala Lys Arg Ser
    355                 360                 365

CAG ACA ACG GGT CAT TTG GTC ATC TGC ATG CCC AAG   1188
Gln Thr Thr Gly His Leu Val Ile Cys Met Pro Lys
                370                 375

GTA GGA GAA GTA ATC ACA GGT GGT CAG CGA GCA TTC   1224
Val Gly Glu Val Ile Thr Gly Gly Gln Arg Ala Phe
        380                 385

AAA TCT ATG AAA ACT ACC TCG GAC AGG AGC AGA GAA   1260
Lys Ser Met Lys Thr Thr Ser Asp Arg Ser Arg Glu
390                 395                 400

CAA ACA AAT ACA AGA AGC AAG CAC ATG GAG AAA CTA   1296
Gln Thr Asn Thr Arg Ser Lys His Met Glu Lys Leu
            405                 410

GAA GTA GAC CCT AGC AAG CAC TCA TTC CCT GAT GTG   1332
Glu Val Asp Pro Ser Lys His Ser Phe Pro Asp Val
    415                 420                 425

ACT AAC ATA GTT CAA GAG AAA AAA CAC ACA CCC AGA   1368
Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg
                430                 435

AGA CGA CCT GAA CCC AAA ATT ATA CCA AGT GAG GAA   1404
Arg Arg Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu
        440                 445

GAC CCA ACC TTT GAA GAC AAC CCT GAA GTG CCT CCG   1440
Asp Pro Thr Phe Glu Asp Asn Pro Glu Val Pro Pro
450                 455                 460

CTG ATT TGA                                      1446
Leu Ile
```

SEQ ID NO:31

CHART C

```
AGCTGGGAGC GCAGAGGCTC ACGCCTGTAA TCCATCATTT                    40

GCTTAGGTCT GATCAATCTG CTCCACACAA TTTCTCAGTG                    80

ATCCTCTGCA TCTCTGCCTA CAAGGGCCTC CCTGACACCC                   120

AAGTTCATAT TGCTCAGAAA CAGTGAACTT GAGTTTTTCG                   160

TTTTACCTTG ATCTCTCTCT GACAAAGAAA TCCAGATGAT                   200

GCAACACCTG ATGAAGACAA TACATGGAAA                              230

                ATG ACA GTC TTG GAA ATA ACT TTG                   254
                Met Thr Val Leu Glu Ile Thr Leu
                                5

GCT GTC ATC CTG ACT CTA CTG GGA CTT GCC ATC CTG                   290
Ala Val Ile Leu Thr Leu Leu Gly Leu Ala Ile Leu
    10                  15                  20

GCT ATT TTG TTA ACA AGA TGG GCA CGA CGT AAG CAA                   326
Ala Ile Leu Leu Thr Arg Trp Ala Arg Arg Lys Gln
                25                  30

AGT GAA ATG TAT ATC TCC AGA TAC AGT TCA GAA CAA                   362
Ser Glu Met Tyr Ile Ser Arg Tyr Ser Ser Glu Gln
        35                  40

AGT GCT AGA CTT CTG GAC TAT GAG GAT GGT AGA GGA                   398
Ser Ala Arg Leu Leu Asp Tyr Glu Asp Gly Arg Gly
45                  50                  55

TCC CGA CAT GCA TAT CAA CAC AAA GTG ACA CTT CAT                   434
Ser Arg His Ala Tyr Gln His Lys Val Thr Leu His
            60                  65

ATG ATA ACC GAG AGA GAT CCA AAA AGA GAT TAC ACA                   470
Met Ile Thr Glu Arg Asp Pro Lys Arg Asp Tyr Thr
    70                  75                  80

CCA TCA ACC AAC TCT CTA GCA CTG TCT CGA TCA AGT                   506
Pro Ser Thr Asn Ser Leu Ala Leu Ser Arg Ser Ser
                85                  90

ATT GCT TTA CCT CAA GGA TCC ATG AGT AGT ATA AAA                   542
Ile Ala Leu Pro Gln Gly Ser Met Ser Ser Ile Lys
        95                  100

TGT TTA CAA ACA ACT GAA GAA CCT CCT TCC AGA ACT                   578
Cys Leu Gln Thr Thr Glu Glu Pro Pro Ser Arg Thr
105                 110                 115
```

GCA GGA GCC ATG ATG CAA TTC ACA GCC CTA TTC CCG 614
Ala Gly Ala Met Met Gln Phe Thr Ala Leu Phe Pro
120 125

GAG CTA CAG GAC CTA TCA AGC TCT CTC AAA AAA CCA 650
Glu Leu Gln Asp Leu Ser Ser Ser Leu Lys Lys Pro
130 135 140

TTG TGC AAA CTC CAG GAC CTA TTG TAC AAT ATC TGG 686
Leu Cys Lys Leu Gln Asp Leu Leu Tyr Asn Ile Trp
145 150

ATC CAA TGT CAG ATC GCA TCT CAC ACA ATC ACT GGT 722
Ile Gln Cys Gln Ile Ala Ser His Thr Ile Thr Gly
155 160

CAC CTT CAG CAC CCG CGG TCA CCC ATG GCA CCC ATA 758
His Leu Gln His Pro Arg Ser Pro Met Ala Pro Ile
165 170 175

ATA ATT TCA CAG AGA ACC GCA AGT CAG CTG GCA GCA 794
Ile Ile Ser Gly Arg Thr Ala Ser Gln Leu Ala Ala
180 185

CCT ATA AGA ATA CCT CAA GTT CAC ACT ATG GAC AGT 830
Pro Ile Arg Ile Pro Gln Val His Thr Met Asp Ser
190 195 200

TCT GGA AAA ATC ACA CTG ACT CCT GTG GTT ATA TTA 866
Ser Gly Lys Ile Thr Leu Thr Pro Val Val Ile Leu
205 210

ACA GGT TAC ATG GAC GAA GAA CTT CGA AAA AAA TCT 902
Thr Gly Tyr Met Asp Glu Glu Leu Arg Lys Lys Ser
215 220

TGT TCC AAA ATC CAG ATT CTA AAA TGT GGA GGC ACT 938
Cys Ser Lys Ile Gln Ile Leu Lys Cys Gly Gly Thr
225 230 235

GCA AGG TCT CAG ATA GCC GAG AAG AAA ACA AGG AAG 974
Ala Arg Ser Gln Ile Ala Glu Lys Lys Thr Arg Lys
240 245

CAA CTA AAG AAT GAC ATC ATA TTT ACG AAT TCT GTA 1010
Gln Leu Lys Asn Asp Ile Ile Phe Thr Asn Ser Val
250 255 260

GAA TCC TTG AAA TCA GCA CAC ATA AAG GAG CCA GAA 1046
Glu Ser Leu Lys Ser Ala His Ile Lys Glu Pro Glu
265 270

AGA GAA GGA AAA GGC ACT GAT TTA GAG AAA GAC AAA 1082
Arg Glu Gly Lys Gly Thr Asp Leu Glu Lys Asp Lys
275 280

ATA GGA ATG GAG GTC AAG GTA GAC AGT GAC GCT GGA 1118
Ile Gly Met Glu Val Lys Val Asp Ser Asp Ala Gly
285 290 295

ATA CCA AAA AGA CAG GAA ACC CAA CTA AAA ATC AGT 1154
Ile Pro Lys Arg Gln Glu Thr Gln Leu Lys Ile Ser
300 305

GAA GAT GAG TAT ACC ACA AGG ACA GGG AGC CCA AAT 1190
Glu Asp Glu Tyr Thr Thr Arg Thr Gly Ser Pro Gln
310 315 320

AAA GAA AAG TGT GTC AGA TGT ACC AAG AGG ACA GGA 1226
Lys Glu Lys Cys Val Arg Cys Thr Lys Arg Thr Gly
325 330

GTC CAA GTA AAG AAG AGT GAG TCA GGT GTC CCA AAA 1262
Val Gln Val Lys Lys Ser Glu Ser Gly Val Pro Lys
335 340

GGA CAA GAA GCC CAA GTA ACG AAG AGT GGG TTG GTT 1298
Gly Gln Glu Ala Gln Val Thr Lys Ser Gly Leu Val
345 350 355

GTA CTG AAA GGA CAG GAA GCC CAG GTA GAG AAG AGT 1334
Val Leu Lys Gly Gln Glu Ala Gln Val Glu Lys Ser
360 365

GAG ATG GGT GTG CCA AGA AGA CAG GAA TCC CAA GTA 1370
Glu Met Gly Val Pro Arg Arg Gln Glu Ser Gln Val
370 375 380

AAG AAG AGT CAG TCT GGT GTC TCA AAG GGA CAG GAA 1406
Lys Lys Ser Gln Ser Gly Val Ser Lys Gly Gln Glu
385 390

GCC CAG GTA AAG AAG AGG GAG TCA GTT GTA CTG AAA 1442
Ala Gln Val Lys Lys Arg Glu Ser Val Val Leu Lys
395 400

GGA CAG GAA GCC CAG GTA GAG AAG AGT GAG TTG AAG 1478
Gly Gln Glu Ala Gln Val Glu Lys Ser Glu Leu Lys
405 410 415

GTA CCA AAA GGA CAA GAA GGC CAA GTA GAG AAG ACT 1514
Val Pro Lys Gly Gln Glu Gly Gln Val Glu Lys Thr
420 425

GAG GCA GAT GTG CCA AAG GAA CAA GAG GTC CAA GAA 1550
Glu Ala Asp Val Pro Lys Glu Gln Glu Val Gln Glu
430 435 440

AAG AAG AGT GAG GCA GGT GTA CTG AAA GGA CCA GAA 1586
Lys Lys Ser Glu Ala Gly Val Leu Lys Gly Pro Glu
445 450

```
TCC CAA GTA AAG AAC ACT GAG GTG AGT GTA CCA GAA                    1622
Ser Gln Val Lys Asn Thr Glu Val Ser Val Pro Glu
        455                 460

ACA CTG GAA TCC CAA GTA AAG AAG AGT GAG TCA GGT                    1658
Thr Leu Glu Ser Gln Val Lys Lys Ser Glu Ser Gly
465                 470                 475

GTA CTA AAA GGA CAG GAA GCC CAA GAA AAG AAG GAG                    1694
Val Leu Lys Gly Gln Glu Ala Gln Glu Lys Lys Glu
            480                 485

AGT TTT GAG GAT AAA GGA AAT AAT GAT AAA GAA AAG                    1730
Ser Phe Glu Asp Lys Gly Asn Asn Asp Lys Glu Lys
    490                 495                 500

GAG AGA GAT GCA GAG AAA GAT CCA AAT AAA AAA GAA                    1766
Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu
                505                 510

AAA GGT GAC AAA AAC ACA AAA GGT GAC AAA GGA AAG                    1802
Lys Gly Asp Lys Asn Thr Lys Gly Asp Lys Gly Lys
        515                 520

GAC AAA GTT AAA GGA AAG AGA GAA TCA GAA ATC AAT                    1838
Asp Lys Val Lys Gly Lys Arg Glu Ser Glu Ile Asn
525                 530                 535

GGT GAA AAA TCA AAA GGC TCG AAA AGG CGA AGG CAA                    1874
Gly Glu Lys Ser Lys Gly Ser Lys Arg Arg Arg Gln
            540                 545

ATA CAG GAA GGA AGT ACA ACA AAA AAG TGG AAG AGT                    1910
Ile Gln Glu Gly Ser Thr Thr Lys Lys Trp Lys Ser
    550                 555                 560

AAG GAT AAA TTT TTT AAA GGC CCA TAA GACAAGTGAT                     1946
Lys Asp Lys Phe Phe Lys Gly Pro
                565

TATTATGATT CCCATACTCC AGATACAAAC CATATCCCAG                        1986

CCATTGCCTA AACAGATTAC AATTATAAAA TCCCTTTCAT                        2026

CTTCATATCA CAGTTTCTGC TCTTCAGAAG TTTCACCCTT                        2066

TTTAATCTCT CAGCCACAAA CCTCAGTTCC AATATTGTTA                        2106

TAAGTTAAGA CGTATATGAT TCCGTCAAGA AAGACTGGAT                        2146

ACTTTCTGAA GTAAAACATT TTAATTAAAG AAAAAAAA                          2184
```

SEQ ID NO:32

SEQUENCE LISTIN

(1) GENERAL INFORMATION:
(i) APPLICANT: Goldberg, Erwin
(i) APPLICANT: O'Hern, Patricia A.
(ii) TITLE OF INVENTION: Proteins And Peptides For Contraceptive Vaccines And Fertility Diagnostics
(iii) NUMBER OF SEQUENCES: 32
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Willian Brinks Hofer Gilson & Lione
(B) STREET: P.O. Box 10395
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: USA
(F) ZIP: 60610
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.50 inch, 2 Mb storage
(B) COMPUTER: IBM XT compatible
(C) OPERATING SYSTEM: MS-DOS
(D) SOFTWARE: WordPerfect 5.1
(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 11-JAN-1996
(C) CLASSIFICATION:
(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Crook, Wannell M.
(B) REGISTRATION NUMBER: 31071
(C) REFERENCE/DOCKET NUMBER: 6793/9
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (312)321-4229
(B) TELEFAX: (312)321-4299
(2) INFORMATION FOR SEQ ID NO:1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:1:

Met Gly Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser
5 10

Pro Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn
15 20 25

Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr Lys
30 35 40

Gln

(2) INFORMATION FOR SEQ ID NO:2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:2:

Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr
5 10

Lys Gln
15

(2) INFORMATION FOR SEQ ID NO:3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:3:

Thr Val Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser
5 10

Arg Thr Lys Gln
15

(2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:4:

Ser Phe Val Thr Val Asn Ala Glu Glu Gln Glu Lys Gln Phe
5 10

Val Ser Ser Arg Thr Lys Gln
15 20

(2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:5:

Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

(2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:6:

Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser Thr Lys
5 10

Ile Tyr Ser Tyr Phe Pro Ser Val
15 20

(2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:7:

Asn Ala Gly Glu Gln Glu Lys Gln Phe Leu Ser Ser Arg Thr
5 10

Lys Gln Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser
15 20 25

Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:8:

Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu Asp Pro Thr Phe 15
20 25

Glu

(2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:9:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys
15

(2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:10:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg Pro Glu
5 10

Pro Lys Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35

(2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:11:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala
15 20

(2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:12:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn
15

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:13:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala Gly Pro Ser Leu Val Asp Asp Ala
15 20 25

Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

(2) INFORMATION FOR SEQ ID NO:14:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:14:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

(2) INFORMATION FOR SEQ ID NO:15:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:15:

Met Asn Pro Thr Glu Thr Lys Ala Ile Pro Val Ser Gln Gln
5 10

Met Glu Gly Pro His Leu Pro Asn Lys Lys Lys His Lys Lys
15 20 25

Gln Ala Val Lys Thr Glu Pro Glu Lys Lys Ser Gln Ser
30 35 40

(2) INFORMATION FOR SEQ ID NO:16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:16:

Lys Gln Gln Glu Ala Gln Val Lys Lys Ser Glu Ser Gly Val
5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO:17:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:17:

Lys Arg Thr Gly Val Gln Val Lys Lys Ser Glu Ser Gly Val
5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO:18:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:18:

Lys Gly Gln Glu Ala Gln Val Thr Lys Ser Gly Leu Val Val
5 10

Leu
15

(2) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:19:

Lys Gly Gln Glu Ala Gln Val Glu Lys Ser Glu Met Gly Val
5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO:20:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:20:

Arg Arg Gln Glu Ser Gln Val Lys Lys Ser Gln Ser Gly Val
5 10

Ser
15

(2) INFORMATION FOR SEQ ID NO:21:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:21:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu
15

(2) INFORMATION FOR SEQ ID NO:22:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:22:

Lys Gly Gln Glu Ala Gln Val Glu Lys Ser Glu Leu Lys Val
5 10

Pro
14

(2) INFORMATION FOR SEQ ID NO:23:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:23:

Lys Gly Gln Glu Gly Gln Val Glu Lys Thr Glu Ala Glu Cys
5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO:24:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:24:

Lys Glu Gln Glu Val Gln Glu Lys Lys Ser Glu Ala Gly Val
5 10

Leu
15

(2) INFORMATION FOR SEQ ID NO:25:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:25:

Lys Gly Pro Glu Phe Gln Val Lys Asn Thr Glu Val Ser Val
5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO:26:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:26:

Glu Thr Leu Glu Ser Gln Val Lys Lys Ser Glu Ser Gly Val
5 10

Leu
15

(2) INFORMATION FOR SEQ ID NO:27:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:27:

Lys Gly Gln Glu Ala Gln Glu Lys Lys Glu Ser Phe Glu Asp
5 10

Lys
15

(2) INFORMATION FOR SEQ ID NO:28:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:28:

AGGAGAACAA CAACC 15

(2) INFORMATION FOR SEQ ID NO:29:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY:
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:29:

Asn Ala Glu Gly Gln Glu Lys Gln Phe Leu Ser Ser Arg Thr
5 10

Lys Gln
15

(2) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1927 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:30:

CTTGATATCG AATTCGGGGGG AGTCTCCCT GACTTCCAGC 40

AACAATCCTT GAGTCTGAGA CTGCCCTGGC CTAAG ATG GGC 81
Met Gly

CAG TTT CTA TCT TCG ACT TTC TTG GAG GGC TCA CCG 117
Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser Pro
5 10

GCC ACA GTG TCG ACG ATA AGC TTT GTG ACG GTG AAC 153
Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn
15 20 25

GCA GAG GAG CAA GAG AAG CAG TTC GTA TCT TCC AGG 189
Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg
30 35

ACC AAG CAA AAA GCT AAA GAA GAA AAA CTA GAG AAG 225
Thr Lys Gln Lys Ala Lys Glu Glu Lys Leu Glu Lys
40 45 50

TGT GGT GAG GAT GAT GAA ACA ATC CCA TCT GAG TAC 261
Cys Gly Glu Asp Asp Glu Thr Ile Pro Ser Glu Tyr
55 60

```
AGA TTA AAA CCA GCC ACG GAT AAA GAT GGA AAA CCA   297
Arg Leu Lys Pro Ala Thr Asp Lys Asp Gly Lys Pro
        65                  70

CTA TTG CCA GAG CCT GAA GAA AAA CCC AAG CCT CGG   333
Leu Leu Pro Glu Pro Glu Glu Lys Pro Lys Pro Arg
75                  80                  85

AGT GAA TCA GAA CTC ATT GAT GAA CTT TCA GAA GAT   369
Ser Glu Ser Glu Leu Ile Asp Glu Leu Ser Glu Asp
            90                  95

TTC GAC CTG TCT GAA TGT AAA GAG AAA CCA TCT AAG   405
Phe Asp Leu Ser Glu Cys Lys Glu Lys Pro Ser Lys
    100                 105                 110

CCA ACT GAA AAG ACA GAA GAA TCT AAG GCC GCT GCT   441
Pro Thr Glu Lys Thr Glu Glu Ser Lys Ala Ala Ala
                115                 120

CCA GCT CCT GTG TCG GAG GCT GTG TCT CGG ACC TCC   477
Pro Ala Pro Val Ser Glu Ala Val Ser Arg Thr Ser
        125                 130

ATG TGT AGT ATA CAG TCA GCA CCC CCT GAG CCG GCT   513
Met Cys Ser Ile Gln Ser Ala Pro Pro Glu Pro Ala
135                 140                 145

ACC TTG AAG GTC ACA GTG CCA GAT GAT GCT GTA GAA   549
Thr Leu Lys Val Thr Val Pro Asp Asp Ala Val Glu
            150                 155

GCC TTG GCT GAT AGC CTG GGG AAA AAG GAA GCA GAT   585
Ala Leu Ala Asp Ser Leu Gly Lys Lys Glu Ala Asp
    160                 165                 170

CCA GAA GAT GGA AAA CCT GTG ATG GAT AAA GCT AAG   621
Pro Glu Asp Gly Lys Pro Val Met Asp Lys Val Lys
                175                 180

GAG AAG GCC AAA GAA GAA GAC CGT GAA AAG CTT GGT   657
Glu Lys Ala Lys Glu Glu Asp Arg Glu Lys Leu Gly
        185                 190

GAA AAA GAA GAA ACA ATT CCT CCT GAT TAT ATA TTA   693
Glu Lys Glu Glu Thr Ile Pro Pro Asp Tyr Ile Leu
195                 200                 205

GAA GAG GTC AAG GAT AAA GAT GGA AAG CCA CTC CTG   729
Glu Glu Val Lys Asp Lys Asp Gly Lys Pro Leu Leu
            210                 215

CCA AAA GAG TCT AAG GAA CAG CTT CCA CCC ATG AGT   765
Pro Lys Glu Ser Lys Glu Gln Leu Pro Pro Met Ser
    220                 225                 230
```

```
GAA GAC TTC CTT CTG GAT GCT TTG TCT GAG GAC TTC   801
Glu Asp Phe Leu Leu Asp Ala Leu Ser Glu Asp Phe
                235                 240

TCT GGT CCA CAA AAT GCT TCA TCT CTT AAA TTT GAA   837
Ser Gly Pro Gln Asn Ala Ser Ser Leu Lys Phe Glu
        240                 245

GAT GCT AAA CTT GCT GCT GCC ATC TCT GAA GTG GTT   873
Asp Ala Lys Leu Ala Ala Ala Ile Ser Glu Val Val
250                 255                 260

TCC CAA ACC CCA GCT TCA ACG ACC CAA GCT GGA GCC   909
Ser Gln Thr Pro Ala Ser Thr Thr Gln Ala Gly Ala
            265                 270

CCA CCC CGT GAT ACC TCG AGT GAC AAA GAC CTC GAT   945
Pro Pro Arg Asp Thr Ser Ser Asp Lys Asp Leu Asp
    275                 280                 285

GAT GCC TTG GAT AAA CTC TCT GAC AGT CTA GGA CAA   981
Asp Ala Leu Asp Lys Leu Ser Asp Ser Leu Gly Gln
                290                 300

AGG CAG CCT GAC CCA GAT GAG AAC AAA CCA ATG GAA  1017
Arg Gln Pro Asp Pro Asp Glu Asn Lys Pro Met Glu
        305                 310

GAT AAA GTA AAG GAA AAA GCT AAA GCT GAA CAT AGA  1053
Asp Lys Val Lys Glu Lys Ala Lys Ala Glu His Arg
315                 320                 325

GAC AAG CTT GGA GAG AGA GAT GAC ACT ATC CCA CCT  1089
Asp Lys Leu Gly Glu Arg Asp Asp Thr Ile Pro Pro
            330                 335

GAA TAC AGA CAT CTC CTG GAT GAT AAT GGA CAG GAC  1125
Glu Tyr Arg His Leu Leu Asp Asp Asn Gly Gln Asp
    340                 345                 350

AAA CCA GTG AAG CCA CCT ACA AAG AAA TCA GAG GAT  1161
Lys Pro Val Lys Pro Pro Thr Lys Lys Ser Glu Asp
                355                 360

TCA AAG AAA CCT GCA GAT GAC CAA GAC CCC ATT GAT  1197
Ser Lys Lys Pro Ala Asp Asp Gln Asp Pro Ile Asp
        365                 370

GCT CTC TCA GGA GAT CTG GAC AGC TGT CCC TCC ACT  1233
Ala Leu Ser Gly Asp Leu Asp Ser Cys Pro Ser Thr
375                 380                 385

ACA GAA ACC TCA CAG AAC ACA GCA AAG GAT AAG TGC  1269
Thr Glu Thr Ser Gln Asn Thr Ala Lys Asp Lys Cys
            390                 395
```

AAG AAG GCT GCT TCC AGC TCC AAA GCA CCT AAG AAT 1305
Lys Lys Ala Ala Ser Ser Ser Lys Ala Pro Lys Asn
400 405 410

GGA GGT AAA GCG AAG GAT TCA GCA AAG ACA ACA GAG 1341
Gly Gly Lys Ala Lys Asp Ser Ala Lys Thr Thr Glu
415 420

GAA ACT TCC AAG CCA AAA GAT GAC TAA AGAAATACAAG 1377
Glu Thr Ser Lys Pro Lys Asp Asp
425 430

TTAAGGTATC TGGTATCTGC ATTTAAAATC TTCAGCTGGT 1417

GGATTGTGAC TTTTGAAGAA CAAAAGGCTT TGGCAACAGA 1457

AAACAATTGT TCTGGGTGAT TTCTAGAATG TTTTTTGTTG 1497

AGTCTCTGAA CATCCTAAAT ATTTGTTTGT TATTCTTTTC 1537

CAGAAAGAAA ATGAATTTGA CTGGTTCACC TGTGTACTGA 1577

GTATTGATAA ACTTCGAATT TTTTAAATTT CCTTCAAGGG 1617

AGAGAAAGCT TATATTGGTT TGTTATTCTT TTCCAGAAAG 1657

AAAATGAATT TGACTGGGTT CACTGTGTTA CTGAGTATTG 1697

ATAAACTTTG AATTTTTGCA ATTGCCTTCA ATTTTTAGAG 1737

GAAAAGCTTT ATATTTGTGT TATTACTTCT TCATCTTACA 1777

GTCATCACAG AACACACTGA GACTTGAATC AAGTCAGCAA 1817

CAGAGCAAAA TAAAGGTTAG ATAAGTCCTT GTGTAGCAAA 1857

TTTCGAGCAT AAGAAATAAA ATCTAATTAA TTCTTAGGGT 1897

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1927

(2) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1446 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:31:

AAAGCGTCAT TCGAGGTCCG GGTCCGGCTT GCGGGGTCAG 40

CGAACTGGAG AGGCGCC ATG GGC TGG ATC ACA 72
Met Gly Trp Ile Thr
5

```
GAA GAT CTT ATT AGA CGG AAT GCT GAA CAC AAC GAC   108
Glu Asp Leu Ile Arg Arg Asn Ala Glu His Asn Asp
                10                  15

TGT GTC ATT TTT TCC CTG GAG GAA CTC TCG TTG CAT   144
Cys Val Ile Phe Ser Leu Glu Glu Leu Ser Leu His
        20                  25

CAG CAA GAA ATA GAA AGA CTA GAA CAC ATT GAT AAA   180
Gln Gln Glu Ile Glu Arg Leu Glu His Ile Asp Lys
30                  35                  40

TGG TGC CGG GAT TTA AAA ATT CTC TAT CTT CAA AAT   216
Trp Cys Arg Asp Leu Lys Ile Leu Tyr Leu Gln Asn
            45                  50

AAT CTT ATT GGG AAA ATT GAA AAT GTT AGC AAA CTC   252
Asn Leu Ile Gly Lys Ile Glu Asn Val Ser Lys Leu
    55                  60                  65

AAG AAA CTT GAA TAT TTG AAT TTA GCT TTA AAC AAC   288
Lys Lys Leu Glu Tyr Leu Asn Leu Ala Leu Asn Asn
                70                  75

ATT GAA AAA ATA GAA AAC TTG GAA GGA TGT GAA GAG   324
Ile Glu Lys Ile Glu Asn Leu Glu Gly Cys Glu Glu
        80                  85

CTG GCA AAA CTT GAC CTG ACT GTG AAT TTC ATT GGA   360
Leu Ala Lys Leu Asp Leu Thr Val Asn Phe Ile Gly
90                  95                  100

GAG CTG AGC AGC ATT AAA AAC TTG CAG CAC AAT ATC   396
Glu Leu Ser Ser Ile Lys Asn Leu Gln His Asn Ile
            105                 110

CAT CTG AAG GAG CTC TTT CTC ATG GGG AAC CCA TGT   432
His Leu Lys Glu Leu Phe Leu Met Gly Asn Pro Cys
    115                 120                 125

GCT TCC TTT GAC CAC TAT AGG GAG TTC GTG GTA GCA   468
Ala Ser Phe Asp His Tyr Arg Glu Phe Val Val Ala
                130                 135

ACT CTT CCA CAA TTA AAG TGG TTG GAT GGT AAA GAA   504
Thr Leu Pro Gln Leu Lys Trp Leu Asp Gly Lys Glu
        140                 145

ATA GAG CCT TCA GAA AGG ATT AAG GCA TTG CAG GAC   540
Ile Glu Pro Ser Glu Arg Ile Lys Ala Leu Gln Asp
150                 155                 160

TAT TCA GTA ATT GAA CCA CAA ATC AGA GAG CAG GAA   576
Tyr Ser Val Ile Glu Pro Gln Ile Arg Glu Gln Glu
            165                 170
```

```
AAA GAT CAC TGT CTT AAA CGA GCC AAA CTC AAG GAA                    612
Lys Asp His Cys Leu Lys Arg Ala Lys Leu Lys Glu
    175                 180                 185

GAG GCT CAG AGG AAA CAC CAA GAA GAG GAT AAA AAT                    648
Glu Ala Gln Arg Lys His Gln Glu Glu Asp Lys Asn
                190                 195

GAA GAC AAG AGA AGT AAC GCA GGC TTT GAT GGA CGT                    684
Glu Asp Lys Arg Ser Asn Ala Gly Phe Asp Gly Arg
        200                 205

TGG TAC ACA GAC ATC AAT GCT ACT CTT TCC TCT TTA                    720
Trp Tyr Thr Asp Ile Asn Ala Thr Leu Ser Ser Leu
210                 215                 220

GAG AGC AAA GAC CAC CTA CAG GCA CCA GAC ATA GAG                    756
Glu Ser Lys Asp His Leu Gln Ala Pro Asp Ile Glu
            225                 230

GAA CAC AAC ACA AAG AAA TTA GAC GAT GAC TTG GAA                    792
Glu His Asn Thr Lys Lys Leu Asp Asp Asp Leu Glu
    235                 240                 245

TTC TGG AAT AAG CCC TGT TTG TTT ACT CCT GAA TCA                    828
Phe Trp Asn Lys Pro Cys Leu Phe Thr Pro Glu Ser
                250                 255

AGA TTG GAA ACT CTT AGA CAC ATG GAA AAA CAA CGG                    864
Arg Leu Glu Thr Leu Arg His Met Glu Lys Gln Arg
        260                 265

AAG AAA CAG GAA AAA TTA AGT GAA AAA AAG AAG AAA                    900
Lys Lys Gln Glu Lys Leu Ser Glu Lys Lys Lys Lys
270                 275                 280

GTG AAA CCA CCC AGG ACT TTG ATC ACT GAA GAT GGG                    936
Val Lys Pro Pro Arg Thr Leu Ile Thr Glu Asp Gly
            285                 290

AAA GCC CTA AAT GTG AAT GAG CCC AAA ATT GAC TTC                    972
Lys Ala Leu Asn Val Asn Glu Pro Lys Ile Asp Phe
    295                 300                 305

TCT TTG AAA GAT AAC GAA AAG CAG ATC ATC CTG GAC                   1008
Ser Leu Lys Asp Asn Glu Lys Gln Ile Ile Leu Asp
                310                 315

CTT GCT GTC TAT AGG TAT ATG GAT ACC TCT TTA ATC                   1044
Leu Ala Val Tyr Arg Tyr Met Asp Thr Ser Leu Ile
        320                 325

GAT GTT GAT GTG CAA CCA ACT TAC GTG CGA GTA ATG                   1080
Asp Val Asp Val Gln Pro Thr Tyr Val Arg Val Met
330                 335                 340
```

```
ATC AAA GGA AAG CCA TTT CAG CTT GTC CTT CCT GCA   1116
Ile Lys Gly Lys Pro Phe Gln Leu Val Leu Pro Ala
            345                 350

GAA GTG AAA CCC GAT AGT AGT TCT GCT AAA AGA TCT   1152
Glu Val Lys Pro Asp Ser Ser Ser Ala Lys Arg Ser
    355                 360                 365

CAG ACA ACG GGT CAT TTG GTC ATC TGC ATG CCC AAG   1188
Gln Thr Thr Gly His Leu Val Ile Cys Met Pro Lys
                370                 375

GTA GGA GAA GTA ATC ACA GGT GGT CAG CGA GCA TTC   1224
Val Gly Glu Val Ile Thr Gly Gly Gln Arg Ala Phe
        380                 385

AAA TCT ATG AAA ACT ACC TCG GAC AGG AGC AGA GAA   1260
Lys Ser Met Lys Thr Thr Ser Asp Arg Ser Arg Glu
390                 395                 400

CAA ACA AAT ACA AGA AGC AAG CAC ATG GAG AAA CTA   1296
Gln Thr Asn Thr Arg Ser Lys His Met Glu Lys Leu
            405                 410

GAA GTA GAC CCT AGC AAG CAC TCA TTC CCT GAT GTG   1332
Glu Val Asp Pro Ser Lys His Ser Phe Pro Asp Val
    415                 420                 425

ACT AAC ATA GTT CAA GAG AAA AAA CAC ACA CCC AGA   1368
Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg
                430                 435

AGA CGA CCT GAA CCC AAA ATT ATA CCA AGT GAG GAA   1404
Arg Arg Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu
        440                 445

GAC CCA ACC TTT GAA GAC AAC CCT GAA GTG CCT CCG   1440
Asp Pro Thr Phe Glu Asp Asn Pro Glu Val Pro Pro
450                 455                 460

CTG ATT TGA                                      1446
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2184 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION:SEQ ID NO:32:

```
AGCTGGGAGC GCAGAGGCTC ACGCCTGTAA TCCATCATTT        40

GCTTAGGTCT GATCAATCTG CTCCACACAA TTTCTCAGTG        80

ATCCTCTGCA TCTCTGCCTA CAAGGGCCTC CCTGACACCC       120
```

```
AAGTTCATAT TGCTCAGAAA CAGTGAACTT GAGTTTTTCG                    160

TTTTACCTTG ATCTCTCTCT GACAAAGAAA TCCAGATGAT                    200

GCAACACCTG ATGAAGACAA TACATGGAAA                               230

                    ATG ACA GTC TTG GAA ATA ACT TTG             254
                    Met Thr Val Leu Glu Ile Thr Leu
                                    5

GCT GTC ATC CTG ACT CTA CTG GGA CTT GCC ATC CTG             290
Ala Val Ile Leu Thr Leu Leu Gly Leu Ala Ile Leu
    10                  15                  20

GCT ATT TTG TTA ACA AGA TGG GCA CGA CGT AAG CAA             326
Ala Ile Leu Leu Thr Arg Trp Ala Arg Arg Lys Gln
                25                  30

AGT GAA ATG TAT ATC TCC AGA TAC AGT TCA GAA CAA             362
Ser Glu Met Tyr Ile Ser Arg Tyr Ser Ser Glu Gln
        35                  40

AGT GCT AGA CTT CTG GAC TAT GAG GAT GGT AGA GGA             398
Ser Ala Arg Leu Leu Asp Tyr Glu Asp Gly Arg Gly
45                  50                  55

TCC CGA CAT GCA TAT CAA CAC AAA GTG ACA CTT CAT             434
Ser Arg His Ala Tyr Gln His Lys Val Thr Leu His
            60                  65

ATG ATA ACC GAG AGA GAT CCA AAA AGA GAT TAC ACA             470
Met Ile Thr Glu Arg Asp Pro Lys Arg Asp Tyr Thr
    70                  75                  80

CCA TCA ACC AAC TCT CTA GCA CTG TCT CGA TCA AGT             506
Pro Ser Thr Asn Ser Leu Ala Leu Ser Arg Ser Ser
                85                  90

ATT GCT TTA CCT CAA GGA TCC ATG AGT AGT ATA AAA             542
Ile Ala Leu Pro Gln Gly Ser Met Ser Ser Ile Lys
        95                  100

TGT TTA CAA ACA ACT GAA GAA CCT CCT TCC AGA ACT             578
Cys Leu Gln Thr Thr Glu Glu Pro Pro Ser Arg Thr
105                 110                 115

GCA GGA GCC ATG ATG CAA TTC ACA GCC CTA TTC CCG             614
Ala Gly Ala Met Met Gln Phe Thr Ala Leu Phe Pro
            120                 125

GAG CTA CAG GAC CTA TCA AGC TCT CTC AAA AAA CCA             650
Glu Leu Gln Asp Leu Ser Ser Ser Leu Lys Lys Pro
    130                 135                 140
```

```
TTG TGC AAA CTC CAG GAC CTA TTG TAC AAT ATC TGG   686
Leu Cys Lys Leu Gln Asp Leu Leu Tyr Asn Ile Trp
                145                 150

ATC CAA TGT CAG ATC GCA TCT CAC ACA ATC ACT GGT   722
Ile Gln Cys Gln Ile Ala Ser His Thr Ile Thr Gly
        155                 160

CAC CTT CAG CAC CCG CGG TCA CCC ATG GCA CCC ATA   758
His Leu Gln His Pro Arg Ser Pro Met Ala Pro Ile
165                 170                 175

ATA ATT TCA CAG AGA ACC GCA AGT CAG CTG GCA GCA   794
Ile Ile Ser Gly Arg Thr Ala Ser Gln Leu Ala Ala
            180                 185

CCT ATA AGA ATA CCT CAA GTT CAC ACT ATG GAC AGT   830
Pro Ile Arg Ile Pro Gln Val His Thr Met Asp Ser
    190                 195                 200

TCT GGA AAA ATC ACA CTG ACT CCT GTG GTT ATA TTA   866
Ser Gly Lys Ile Thr Leu Thr Pro Val Val Ile Leu
                205                 210

ACA GGT TAC ATG GAC GAA GAA CTT CGA AAA AAA TCT   902
Thr Gly Tyr Met Asp Glu Glu Leu Arg Lys Lys Ser
        215                 220

TGT TCC AAA ATC CAG ATT CTA AAA TGT GGA GGC ACT   938
Cys Ser Lys Ile Gln Ile Leu Lys Cys Gly Gly Thr
225                 230                 235

GCA AGG TCT CAG ATA GCC GAG AAG AAA ACA AGG AAG   974
Ala Arg Ser Gln Ile Ala Glu Lys Lys Thr Arg Lys
            240                 245

CAA CTA AAG AAT GAC ATC ATA TTT ACG AAT TCT GTA   1010
Gln Leu Lys Asn Asp Ile Ile Phe Thr Asn Ser Val
    250                 255                 260

GAA TCC TTG AAA TCA GCA CAC ATA AAG GAG CCA GAA   1046
Glu Ser Leu Lys Ser Ala His Ile Lys Glu Pro Glu
                265                 270

AGA GAA GGA AAA GGC ACT GAT TTA GAG AAA GAC AAA   1082
Arg Glu Gly Lys Gly Thr Asp Leu Glu Lys Asp Lys
        275                 280

ATA GGA ATG GAG GTC AAG GTA GAC AGT GAC GCT GGA   1118
Ile Gly Met Glu Val Lys Val Asp Ser Asp Ala Gly
285                 290                 295

ATA CCA AAA AGA CAG GAA ACC CAA CTA AAA ATC AGT   1154
Ile Pro Lys Arg Gln Glu Thr Gln Leu Lys Ile Ser
            300                 305
```

GAA GAT GAG TAT ACC ACA AGG ACA GGG AGC CCA AAT 1190
Glu Asp Glu Tyr Thr Thr Arg Thr Gly S r Pro Gln
310 315 320

AAA GAA AAG TGT GTC AGA TGT ACC AAG AGG ACA GGA 1226
Lys Glu Lys Cys Val Arg Cys Thr Lys Arg Thr Gly
325 330

GTC CAA GTA AAG AAG AGT GAG TCA GGT GTC CCA AAA 1262
Val Gln Val Lys Lys Ser Glu Ser Gly Val Pro Lys
335 340

GGA CAA GAA GCC CAA GTA ACG AAG AGT GGG TTG GTT 1298
Gly Gln Glu Ala Gln Val Thr Lys Ser Gly Leu Val
345 350 355

GTA CTG AAA GGA CAG GAA GCC CAG GTA GAG AAG AGT 1334
Val Leu Lys Gly Gln Glu Ala Gln Val Glu Lys Ser
360 365

GAG ATG GGT GTG CCA AGA AGA CAG GAA TCC CAA GTA 1370
Glu Met Gly Val Pro Arg Arg Gln Glu Ser Gln Val
370 375 380

AAG AAG AGT CAG TCT GGT GTC TCA AAG GGA CAG GAA 1406
Lys Lys Ser Gln Ser Gly Val Ser Lys Gly Gln Glu
385 390

GCC CAG GTA AAG AAG AGG GAG TCA GTT GTA CTG AAA 1442
Ala Gln Val Lys Lys Arg Glu Ser Val Val Leu Lys
395 400

GGA CAG GAA GCC CAG GTA GAG AAG AGT GAG TTG AAG 1478
Gly Gln Glu Ala Gln Val Glu Lys Ser Glu Leu Lys
405 410 415

GTA CCA AAA GGA CAA GAA GGC CAA GTA GAG AAG ACT 1514
Val Pro Lys Gly Gln Glu Gly Gln Val Glu Lys Thr
420 425

GAG GCA GAT GTG CCA AAG GAA CAA GAG GTC CAA GAA 1550
Glu Ala Asp Val Pro Lys Glu Gln Glu Val Gln Glu
430 435 440

AAG AAG AGT GAG GCA GGT GTA CTG AAA GGA CCA GAA 1586
Lys Lys Ser Glu Ala Gly Val Leu Lys Gly Pro Glu
445 450

TCC CAA GTA AAG AAC ACT GAG GTG AGT GTA CCA GAA 1622
Ser Gln Val Lys Asn Thr Glu Val Ser Val Pro Glu
455 460

ACA CTG GAA TCC CAA GTA AAG AAG AGT GAG TCA GGT 1658
Thr Leu Glu Ser Gln Val Lys Lys Ser Glu Ser Gly
465 470 475

GTA CTA AAA GGA CAG GAA GCC CAA GAA AAG AAG GAG 1694
Val Leu Lys Gly Gln Glu Ala Gln Glu Lys Lys Glu
480 485

AGT TTT GAG GAT AAA GGA AAT AAT GAT AAA GAA AAG 1730
Ser Phe Glu Asp Lys Gly Asn Asn Asp Lys Glu Lys
490 495 500

GAG AGA GAT GCA GAG AAA GAT CCA AAT AAA AAA GAA 1766
Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu
505 510

AAA GGT GAC AAA AAC ACA AAA GGT GAC AAA GGA AAG 1802
Lys Gly Asp Lys Asn Thr Lys Gly Asp Lys Gly Lys
515 520

GAC AAA GTT AAA GGA AAG AGA GAA TCA GAA ATC AAT 1838
Asp Lys Val Lys Gly Lys Arg Glu Ser Glu Ile Asn
525 530 535

GGT GAA AAA TCA AAA GGC TCG AAA AGG CGA AGG CAA 1874
Gly Glu Lys Ser Lys Gly Ser Lys Arg Arg Arg Gln
540 545

ATA CAG GAA GGA AGT ACA ACA AAA AAG TGG AAG AGT 1910
Ile Gln Glu Gly Ser Thr Thr Lys Lys Trp Lys Ser
550 555 560

AAG GAT AAA TTT TTT AAA GGC CCA TAA GACAAGTGAT 1946
Lys Asp Lys Phe Phe Lys Gly Pro
565

TATTATGATT CCCATACTCC AGATACAAAC CATATCCCAG 1986

CCATTGCCTA AACAGATTAC AATTATAAAA TCCCTTTCAT 2026

CTTCATATCA CAGTTTCTGC TCTTCAGAAG TTTCACCCTT 2066

TTTAATCTCT CAGCCACAAA CCTCAGTTCC AATATTGTTA 2106

TAAGTTAAGA CGTATATGAT TCCGTCAAGA AAGACTGGAT 2146

ACTTTCTGAA GTAAAACATT TTAATTAAAG AAAAAAAA 2184

WE CLAIM:

1. A purified protein which is a testis-specific isoform of calpastatin.

2. The protein of Claim 1 which has the following sequence at its N-terminal:

Met Gly Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser
5 10

Pro Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn
15 20 25

Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr Lys
30 35 40

Gln

SEQ ID NO:1.

3. A peptide capable of producing an antibody that reacts specifically with a testis-specific isoform of calpastatin, said peptide having a sequence comprising a sequence which forms a B-cell epitope found on the testis-specific isoform of calpastatin and not on somatic isoforms of calpastatin.

4. The peptide of Claim 3 having the following sequence:

Met Gly Gln Phe Leu Ser Ser Thr Phe Leu Glu Gly Ser Pro
5 10

Ala Thr Val Ser Thr Ile Ser Phe Val Thr Val Asn Ala Glu
15 20 25

Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr Lys Gln,
30 35 40

SEQ ID NO:1

or a portion thereof that includes the sequence from amino acid 26 through amino acid 41.

5. The peptide of Claim 4 which has the following sequence:

Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser Arg Thr
5 10

Lys Gln
15

SEQ ID NO:2.

6. The peptide of Claim 4 which has the following sequence:

Thr Val Asn Ala Glu Glu Gln Glu Lys Gln Phe Val Ser Ser
5 10

Arg Thr Lys Gln
15

SEQ ID NO:3.

7. The peptide of Claim 4 which has the following sequence:

Ser Phe Val Thr Val Asn Ala Glu Glu Gln Glu Lys Gln Phe
5 10

Val Ser Ser Arg Thr Lys Gln
15 20

SEQ ID NO:4.

8. A peptide having a sequence which comprises th sequence of a T-cell epitope found on a testis-specific isoform of calpastatin.

9. An immunogen comprising the peptide of any one of Claims 3-7 linked to a carrier.

10. The immunogen of Claim 9 wherein the carrier is a peptide having a sequence comprising the sequence of a promiscuous T-cell epitope.

11. The immunogen of Claim 10 wherein the T-cell epitope has the following sequence:

Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

SEQ ID NO:5.

12. The immunogen of Claim 11 wherein the carrier has the following sequence:

Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser Thr Lys
5 10

Ile Tyr Ser Tyr Phe Pro Ser Val
15 20

SEQ ID NO:6.

13. The immunogen of Claim 12 which has the following sequence:

Asn Ala Gly Glu Gln Glu Lys Gln Phe Leu Ser Ser Arg Thr
5 10

Lys Gln Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser
15 20 25

Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35

SEQ ID NO:7.

14. A purified protein which is the protein produced by clone C-2 or a protein at least 70% homologous to the protein produced by clone C-2.

15. The protein of Claim 14 which contains the following sequence:

Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu Asp Pro Thr Phe 15
20 25

Glu

SEQ ID NO:8.

16. A peptide capable of producing an antibody that reacts specifically with the protein of Claim 14, said peptide having a sequence comprising a sequence which forms a B-cell epitope of the protein of Claim 14.

17. The peptide of Claim 16 having the following sequence:

Thr Asn Ile Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys Ile Ile Pro Ser Glu Glu Asp Pro Thr Phe
15 20 25

Glu,

SEQ ID NO:8

or a portion thereof that includes the sequence from amino acid 4 through amino acid 17.

18. The peptide of Claim 17 having the following sequence:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg
5 10

Pro Glu Pro Lys
15

SEQ ID NO:9.

19. A peptide having a sequence which comprises the sequence of a T-cell epitope of the protein of Claim 14.

20. An immunogen comprising the peptide of any one of Claims 15-18 linked to a carrier.

21. The immunogen of Claim 20 wh rein the carrier is a peptide having a sequ nce comprising th sequence of a promiscuous T-cell epitope.

22. The immunogen of Claim 21 wherein the T-cell epitope has the following sequence:

Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

SEQ ID NO:5.

23. The immunogen of Claim 22 wherein the carrier has the following sequence:

Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser Thr Lys
5 10

Ile Tyr Ser Tyr Phe Pro Ser Val
15 20

SEQ ID NO:6.

24. The immunogen of Claim 23 which has the following sequence:

Val Gln Glu Lys Lys His Thr Pro Arg Arg Arg Pro Glu
5 10

Pro Lys Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35

SEQ ID NO:10.

25. A purified protein which is the protein produced by clone L-7 or a protein at least 70% homologous to the protein produced by clone L-7.

26. The protein of Claim 25 which contains th following sequence:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala
15 20

SEQ ID NO:11

and the following sequence:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn
15

SEQ ID NO:12.

27. A peptide capable of producing an antibody that reacts specifically with the protein of Claim 24, said peptide having a sequence comprising a sequence which forms a B-cell epitope of the protein of Claim 24.

28. The peptide of Claim 27 having the following sequence:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala
15 20

SEQ ID NO:11.

29. The peptide of Claim 27 having the following sequence:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn
15

SEQ ID NO:12.

30. A peptide having a sequence which comprises the sequence of a T-cell epitope of the protein of Claim 24.

31. An immunogen comprising the peptide of any one of Claims 26-29 linked to a carrier.

32. The immunogen of Claim 31 wherein the carrier is a peptide having a sequence comprising the sequence of a promiscuous T-cell epitope.

33. The immunogen of Claim 32 wherein the T-cell epitope has the following sequence:

Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr
5 10

Phe Pro Ser Val
15

SEQ ID NO:5.

34. The immunogen of Claim 33 wherein the carrier has the following sequence:

Gly Pro Ser Leu Val Asp Asp Ala Leu Ile Asn Ser Thr Lys
5 10

Ile Tyr Ser Tyr Phe Pro Ser Val
15 20

SEQ ID NO:6.

35. The immunogen of Claim 34 which has the following sequence:

Lys Gly Gln Glu Ala Gln Val Lys Lys Arg Glu Ser Val Val
5 10

Leu Lys Gly Gln Glu Ala Gly Pro Ser Leu Val Asp Asp Ala
15 20 25

Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

SEQ ID NO:13.

36. The immunogen of Claim 34 which has the following sequence:

Lys Glu Arg Asp Ala Glu Lys Asp Pro Asn Lys Lys Glu Lys
5 10

Gly Asp Lys Asn Gly Pro Ser Leu Val Asp Asp Ala Leu Ile
15 20 25

Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val
30 35 40

SEQ ID NO:14.

37. A vaccine comprising a protein of any one of Claims 1-2, 14-15 and 25-26, or an immunogenic portion thereof, in a delivery system.

38. A vaccine comprising a peptide of any one of Claims 3-8, 16-19 and 27-30 in a delivery system.

39. A vaccine comprising an immunogen of Claim 9 in a delivery system.

40. A vaccine comprising an immunogen of Claim 20 in a delivery system.

41. A vaccine comprising an immunogen of Claim 31 in a delivery system.

42. A method of inhibiting fertilization of an egg by sperm comprising administering an effective amount of the vaccine of Claim 37 to a male or female mammal.

43. A method of inhibiting fertilization of an egg by sperm comprising administering an effective amount of the vaccine of Claim 38 to a male or female mammal.

44. A method of inhibiting fertilization of an egg by sperm comprising administering an effective amount of the vaccine of Claim 39 to a male or female mammal.

45. A method of inhibiting fertilization of an egg by sperm comprising administering an effective amount of the vaccine of Claim 40 to a male or female mammal.

46. A method of inhibiting fertilization of an egg by sperm comprising administering an effective amount of the vaccine of Claim 41 to a male or female mammal.

47. An assay for assessing infertility in a patient comprising:

(a) providing one or more of the following:
- (i) a protein of Claim 1;
- (ii) a protein of Claim 14;
- (iii) a protein of Claim 25;
- (iv) a peptide of Claim 3;
- (v) a peptide of Claim 16;
- (vi) a peptide of Claim 27;
- (v) a peptide of Claim 3 linked to a carrier;
- (vi) a peptide of Claim 16 linked to a carrier;
- (vii) a peptide of Claim 27 linked to a carrier;

(b) contacting the protein, peptide or peptide linked to a carrier with a body fluid of the patient; and

(c) determining if the body fluid of the patient contains antibodies that bind to the protein, peptide or peptide linked to a carrier.

48. An assay for assessing infertility in a patient comprising:

(a) providing one or more of the following:

(i) a protein of Claim 2;
(ii) a protein of Claim 15;
(iii) a protein of Claim 26;
(iv) a peptide of Claim 4;
(v) a peptide of Claim 17;
(vi) a peptide of Claim 28;
(vii) a peptide of Claim 29;
(viii) a peptide of Claim 4 linked to a carrier;
(ix) a peptide of Claim 17 linked to a carrier;
(x) a peptide of Claim 28 linked to a carrier;
(xi) a peptide of Claim 29 linked to a carrier;

(b) contacting the protein, peptide or peptide linked to a carrier with a body fluid of the patient; and

(c) determining if the body fluid of the patient contains antibodies that bind to the protein, peptide or peptide linked to a carrier.

49. An kit comprising at least one container, said container containing one or more of the following:

(i) a protein of Claim 1;
(ii) a protein of Claim 14;
(iii) a protein of Claim 25;
(iv) a peptide of Claim 3;
(v) a peptide of Claim 16;
(vi) a peptide of Claim 27;
(v) a peptide of Claim 3 linked to a carrier;
(vi) a peptide of Claim 16 linked to a carrier;
(vii) a peptide of Claim 27 linked to a carrier.

50. An kit comprising at least one container, said container containing one or more of the following:

(i) a protein of Claim 2;
(ii) a protein of Claim 15;
(iii) a protein of Claim 26;
(iv) a peptide of Claim 4;
(v) a peptide of Claim 17;
(vi) a peptide of Claim 28;
(vii) a peptide of Claim 29;
(viii) a peptide of Claim 4 linked to a carrier;
(ix) a peptide of Claim 17 linked to a carrier;
(x) a peptide of Claim 28 linked to a carrier;
(xi) a peptide of Claim 29 linked to a carrier.

51. An isolated DNA molecule coding for the protein of Claim 1, 14 or 25.

52. The DNA molecule of Claim 51 operatively linked to expression control sequences.

53. A host cell comprising the DNA molecule of Claim 51 operatively linked to expression control sequences.

54. A method of producing a protein comprising culturing the host cell of Claim 53 under conditions permitting expression of the protein.

55. A DNA molecule coding for the peptide of Claim 3, 16 or 17.

56. The DNA molecule of Claim 55 wherein the peptide sequence further comprises the sequence of a promiscuous T-cell epitope.

57. The DNA molecule f Claim 55 or 56 operatively linked to expression control sequences.

58. A host cell comprising the DNA molecule of Claim 55 operatively linked to expression control sequences.

59. A method of producing a peptide comprising culturing the host cell of Claim 58 under conditions permitting expression of the peptide.

FIG.1

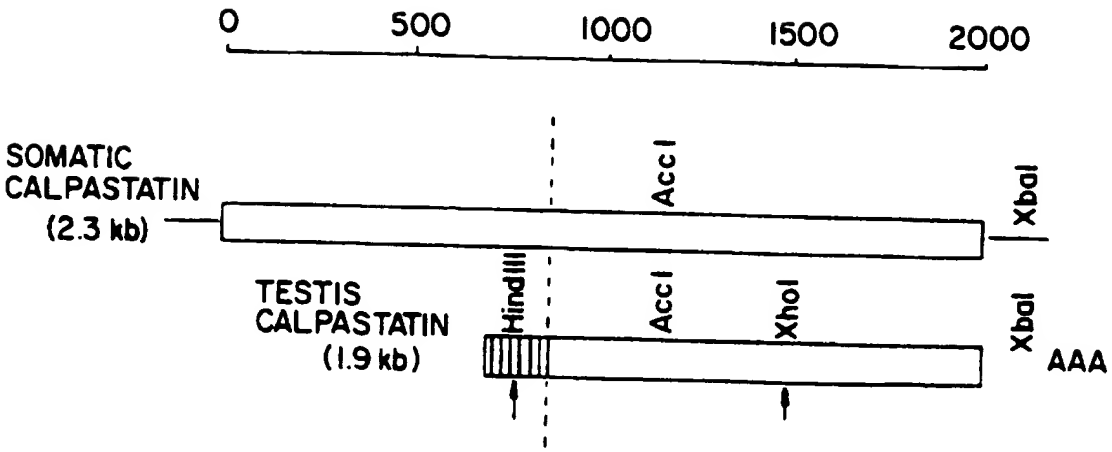

0
500
1000
1500
2000
SOMATIC
CALPASTATIN
(2.3 kb)
Acc I
Xba I
TESTIS
CALPASTATIN
(1.9 kb)
Hind III
Acc I
Xho I
Xba I
AAA

FIG.2

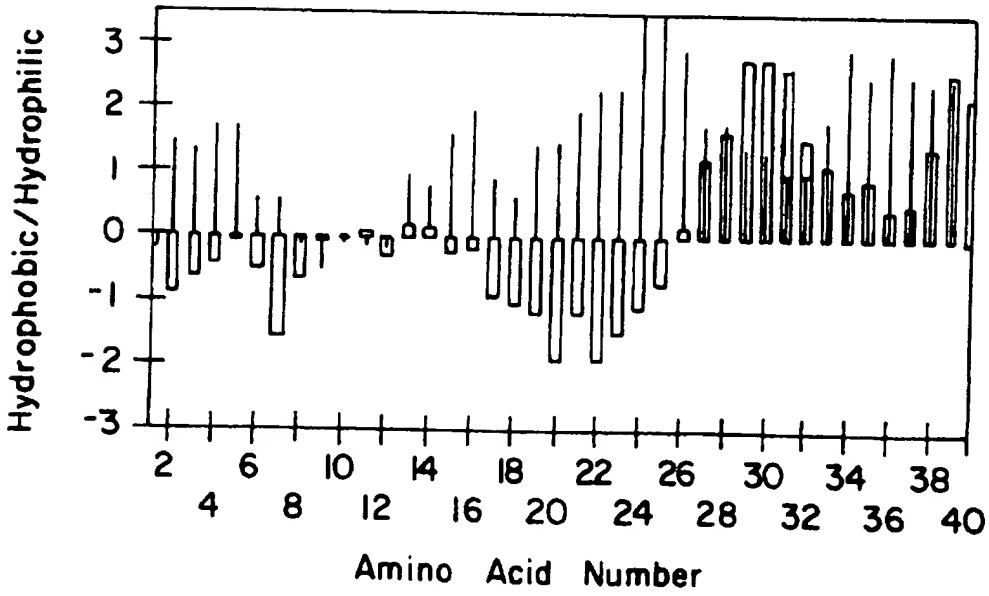

Hydrophobic/Hydrophilic
3
2
1
0
-1
-2
-3
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
Amino Acid Number

FIG.3

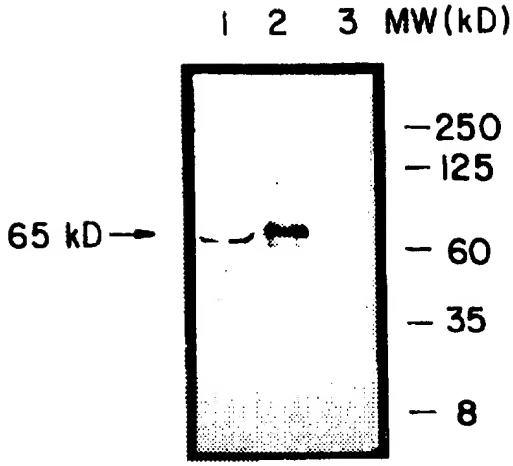

1 2 3 MW(kD)
65 kD
250
125
60
35
8

FIG.4

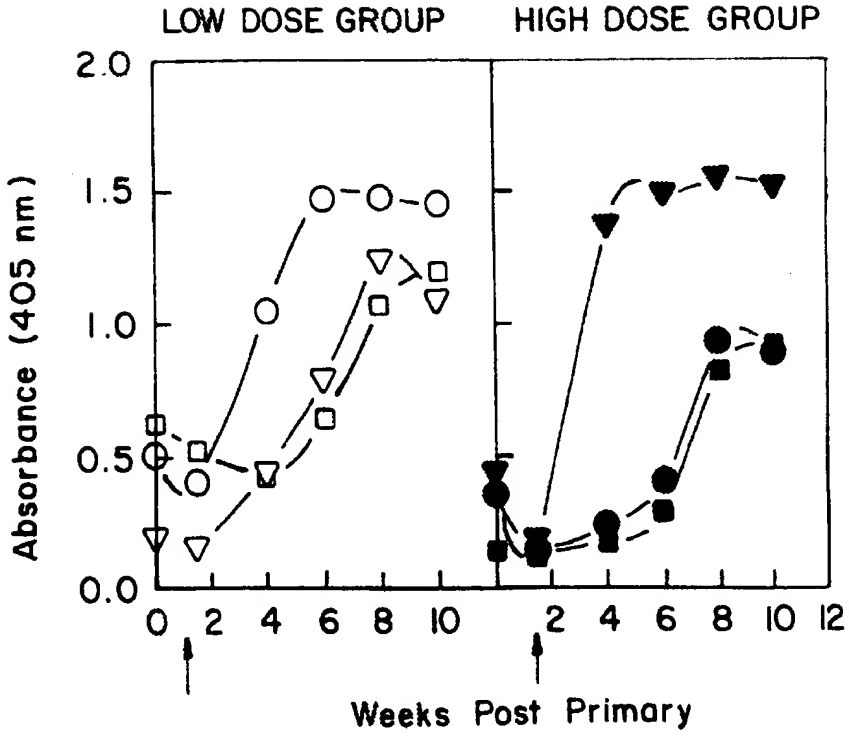

LOW DOSE GROUP
HIGH DOSE GROUP
Absorbance (405 nm)
2.0
1.5
1.0
0.5
0.0
0 2 4 6 8 10
2 4 6 8 10 12
Weeks Post Primary

FIG. 5

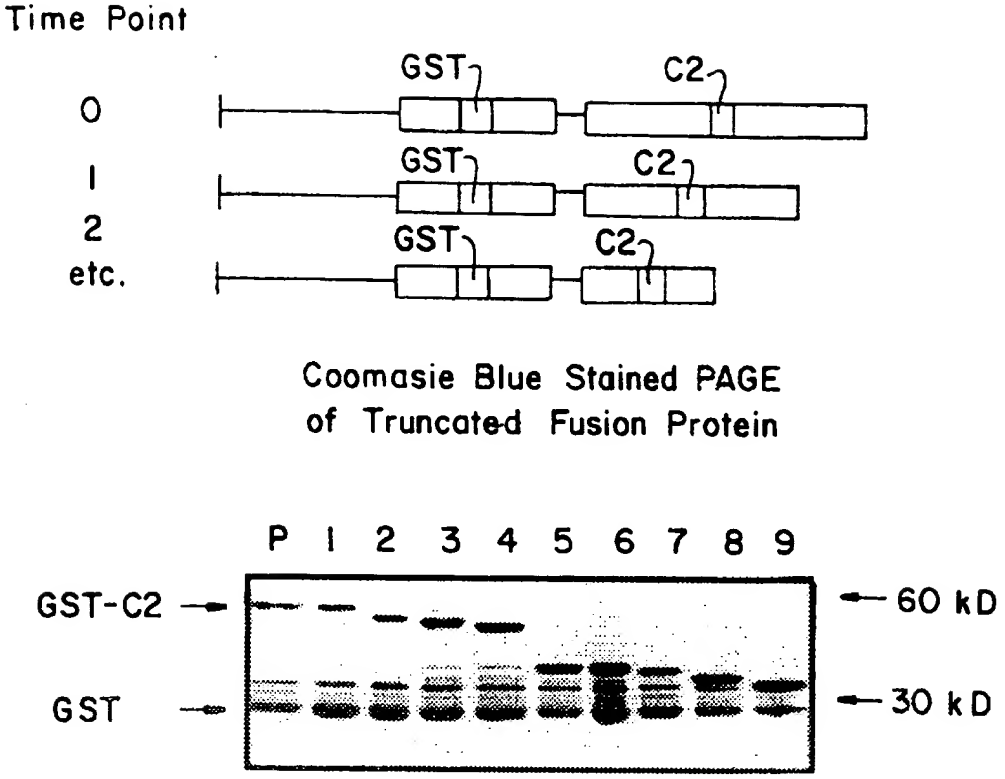

Time Point
0
1
2
etc.
GST
C2
Coomasie Blue Stained PAGE
of Truncated Fusion Protein
P 1 2 3 4 5 6 7 8 9
GST-C2
GST
60 kD
30 kD

FIG. 6

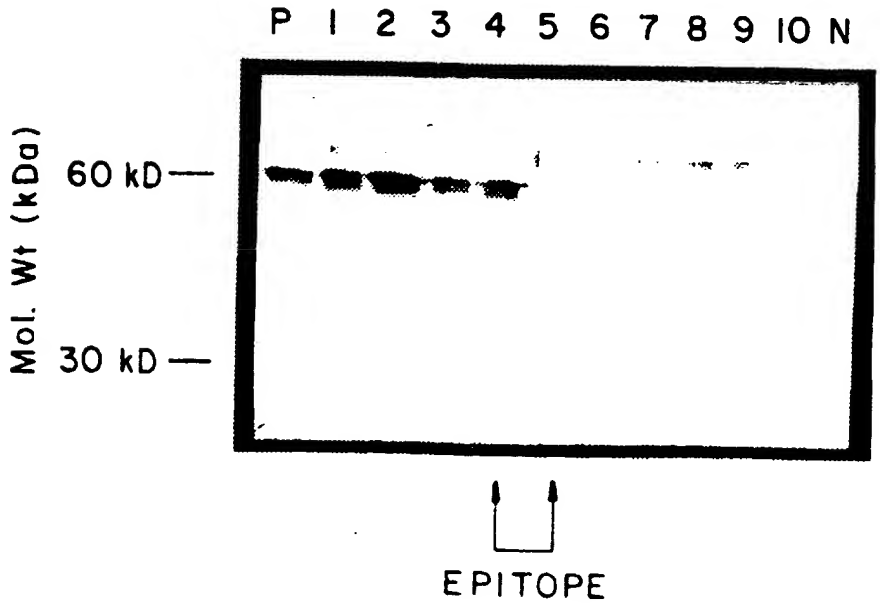

P 1 2 3 4 5 6 7 8 9 10 N
Mol. Wt (kDa)
60 kD
30 kD
EPITOPE

FIG. 7

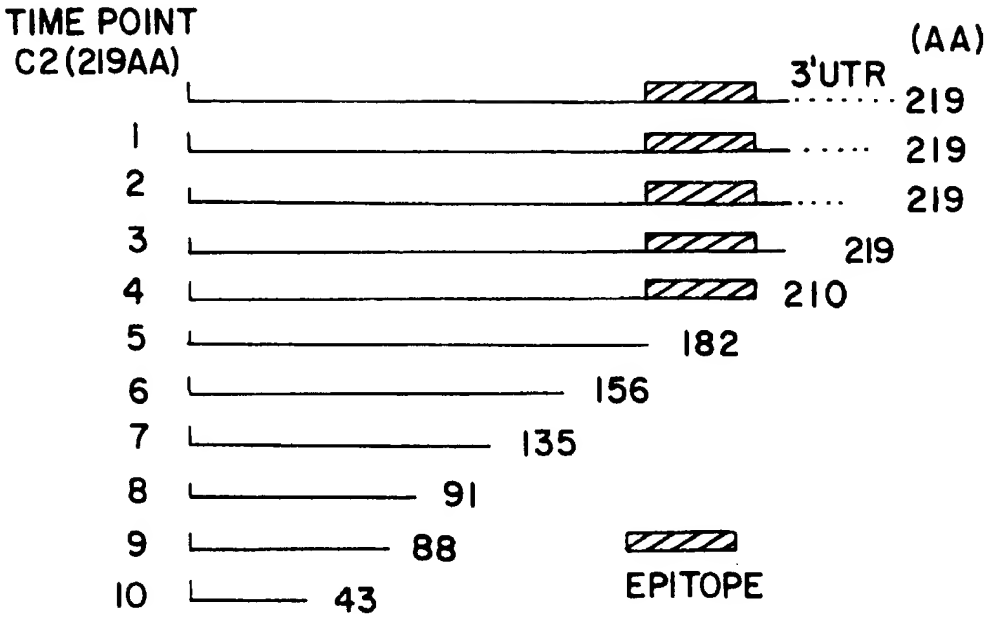

TIME POINT
C2(219AA)
(AA)
3'UTR
219
1
219
2
219
3
219
4
210
5
182
6
156
7
135
8
91
9
88
10
43
EPITOPE

FIG. 8

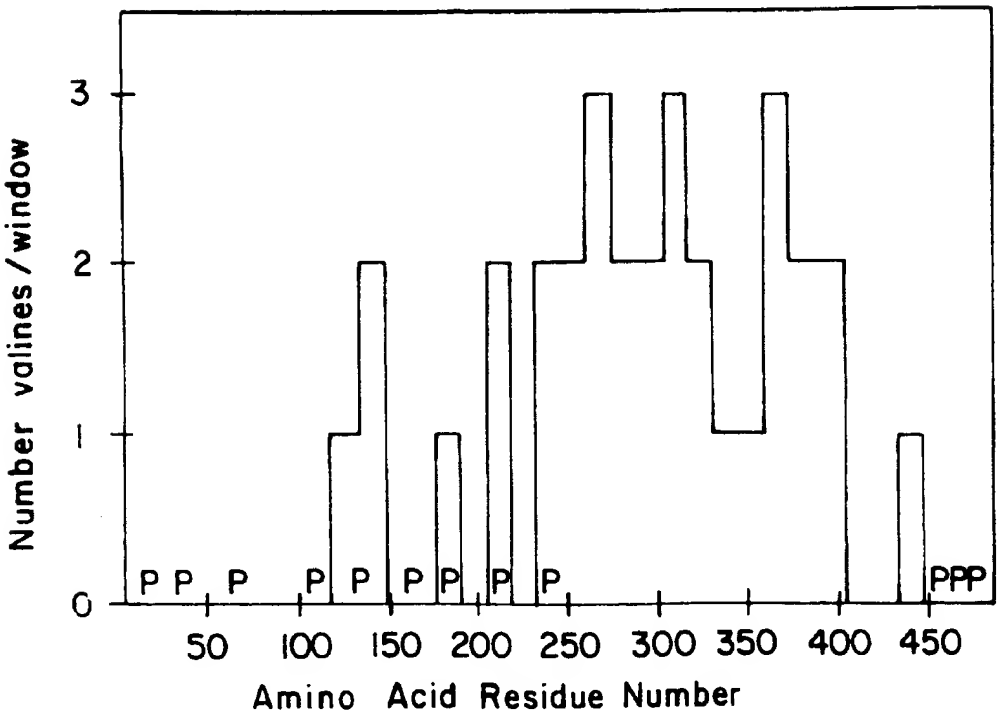

Number valines / window
0
1
2
3
P P P P P P P P P P
PPP
50
100
150
200
250
300
350
400
450
Amino Acid Residue Number

FIG. 9

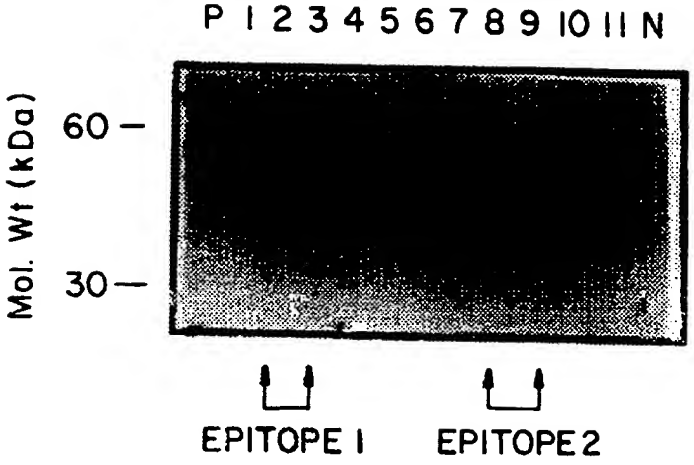
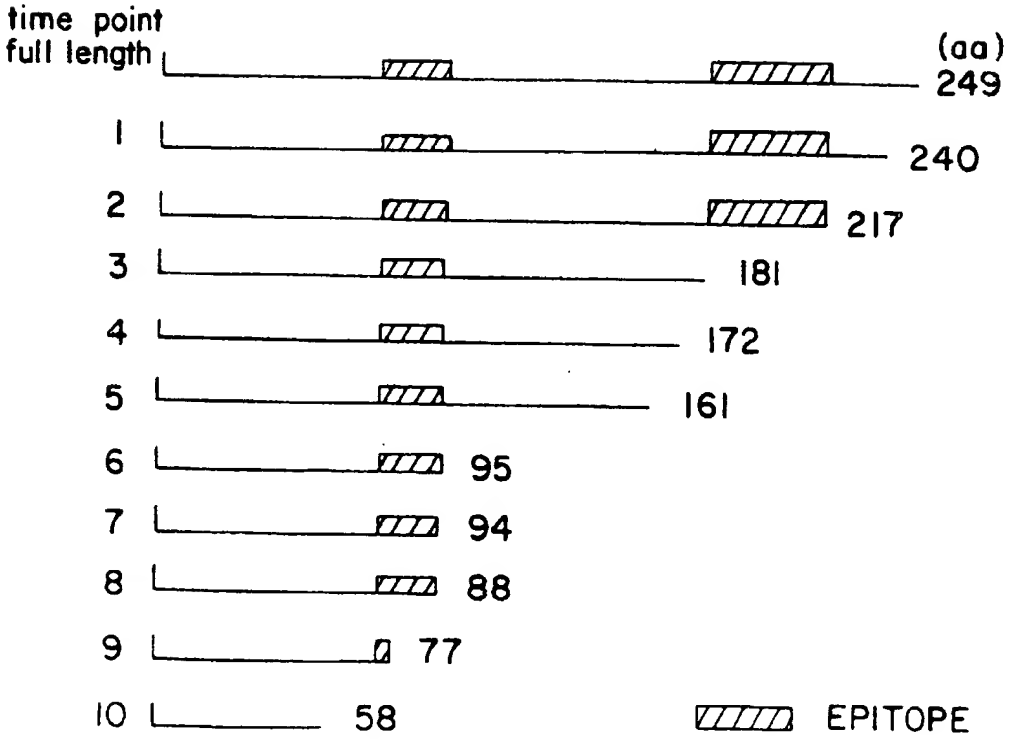

P 1 2 3 4 5 6 7 8 9 10 11 N
Mol. Wt (kDa)
60
30
EPITOPE 1
EPITOPE 2

FIG. 10


time point
full length
(aa)
249
1
240
2
217
3
181
4
172
5
161
6
95
7
94
8
88
9
77
10
58
EPITOPE

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/00908

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A61K 38/10, 38/17; C07K 7/08, 14/81; C12N 1/15, 1/21, 5/10, 15/15

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/184.1, 185.1, 190.1; 435/69.2, 325, 252.3, 254.2; 530/326, 350, 403; 536/23.51

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WANG et al. Calpastatin in human testis. Biochemistry and Molecular Biology International. May 1994, Vol. 33, No. 2, pages 245-252, see abstract.	1, 3, 8, 49, 51-52, 55, 57 ---------- 9-10, 37-39, 42-44,53-54, 56, 58-59
X --- Y	LIANG et al. Human testis cDNAs identified by sera from infertile patients: a molecular biological approach to immunocontraceptive development. Reproduction Fertility and Development. 1994, Vol. 6, pages 297-305, see abstract and page 303, column 2.	1, 3, 8-9, 37-39, 42-44, 49, 51-55, 57-59 ---------- 10, 56

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 24 APRIL 1997	Date of mailing of the international search report 07 MAY 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DAVID SAUNDERS Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/00908

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KAUMAYA et al. Peptide vaccines incorporating a 'promiscuous' T-cell epitope bypass certain haplotype restricted immune responses and provide broad spectrum immunogenicity. Journal of Molecular Recognition. 1993, Vol. 6, pages 81-94, see abstract and Figure 1.	10, 56
Y	O'HEARN et al. The use of Molecular Modelling to Delineate B-cell and T-cell epitopes of human sperm-specific LDH-C_4. Techniques in Protein Chemistry. 1993, Vol. IV, pages 481-490, see pages 481-482 and 488-489.	9, 39, 44

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/00908

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/184.1, 185.1, 190.1; 435/69.2, 325, 252.3, 254.2; 530/326, 324, 403; 536/23.51

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

BIOSIS SEARCH: calpastatin or calpain(w)inhibit?
and
testes or testi? or sperm?

SEQUENCE SEARCHES: Swiss-Prot34, Pir50, Genseq25

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING:
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1.

Group I, claims 1-13, 37-39, 42-44, and 49-59, drawn to calpastatin proteins/peptides, vaccination therewith and production thereof.

Group II, claims 14-24, 37-38, 40, 42-43, 45, and 49-59, drawn to C-2 proteins/ polypeptides, vaccination therewith and production thereof.

Group III, claims 25-38, 41-43, 46, and 49-59, drawn to L-7 proteins/polypeptides, vaccination therewith and production thereof.

Group IV, claims 1-8 and 47-48, drawn to immunoassays using calpastatin.

Group V, claims 14-19 and 47-48, drawn to immunoassays using C-2.

Group VI, claims 25-30 and 47-48, drawn to immunoassays using L-7.

Note the above listing of claims in Group III assumes that applicant intends claims 27 and 30 to depend from claim 25, not 24.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VI do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the calpastatin, C-2 and L-6 proteins of Groups I-III are distinct proteins with no common core structure. They have different sequences that require different searches. They induce the formation of different antibodies when used in a vaccination method. They detect different antibodies when used in an assay. They are produced by different host cells. These Groups thus lack a single inventive concept providing for unity of invention.

Note that in Groups I-III, claims 37-38, 42-43 and 49-59 are listed in common. This is because of the complex dependency of these claims from protein/peptide composition claims pertaining to various of Groups I-III. Claims 37-38, 42-43 and 49-59 will only be examined for the embodiment(s) pertaining to the first recited and additional Groups paid for.

Note that Groups I-III each include the corresponding vaccination methods recited in claims 42-46. Vaccination methods are the first recited use of the protein/peptide compositions and hence included in the unity of invention for each protein/peptide Group. Immunoassay method claims 47-48 of Groups IV-VI are not included in the unity of invention because only one use of one product is permitted in the unity of invention.

Claims 47-48 are listed with each of Groups IV-VI, due to their complex dependencies form protein/peptide composition claims of Groups I-III. Claims 47-48 will only be examined for the embodiment(s) of the paid for extra Groups.